GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 40.1808 Seconds
(without alignments)
3246.633 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5183	100.0	979	19	AAW40309	Human ITAK protein
2	5180	99.9	979	21	AAY95294	Human GEF containi
3	4122	79.5	836	22	AAB65703	Novel protein kina
4	2489	48.0	522	19	AAW69341	Secreted protein o
5	1347	26.0	261	22	ABB11588	Human secreted pro
6	1084	20.9	692	23	AAO15418	Novel human kinase
7	1081	20.9	692	23	AAE22160	Human protein kina
8	848	16.4	841	22	ABB65338	Drosophila melanog
9	731	14.1	146	22	AAU01056	Human secreted pro
10	718	13.9	146	22	AAU01090	Gene 17 Human secr
11	599.5	11.6	1214	22	AAM39211	Human polypeptide
12	599.5	11.6	1214	22	AAU07102	Human novel human
13	599.5	11.6	1242	22	AAM39210	Human polypeptide
14	558	10.8	460	23	AAE21728	Human PKIN-23 prot
15	555	10.7	698	22	AAB65702	Novel protein kina
16	548.5	10.6	330	22	AAB85038	Human SER4 protein
17	547.5	10.6	106	21	AAG02373	Human secreted pro
18	535.5	10.3	649	22	AAU03545	Human protein kina
19	535.5	10.3	654	23	AAU77929	Amino acid sequenc
20	535.5	10.3	683	23	AAU77928	Amino acid sequenc
21	518	10.0	484	23	AAE21727	Human PKIN-22 prot
22	516	10.0	506	22	AAM78344	Human protein SEQ
23	514	9.9	506	23	AAE24136	Human kinase (PKIN
24	510	9.8	527	22	AAM79328	Human protein SEQ
25	510	9.8	527	23	ABB97224	Novel human protei
26	502.5	9.7	345	21	AAY76753	Human protein kina
27	502.5	9.7	345	22	AAE06211	Human protein kina
28	499	9.6	336	21	AAY59148	Human NEK2 protein
29	499	9.6	445	21	AAY92330	Human NlK1 protein
30	476	9.2	360	21	AAY59147	E. nidulans NIMA p
31	465.5	9.0	510	21	AAY68778	Amino acid sequenc
32	447.5	8.6	735	22	ABB66233	Drosophila melanog
33	441	8.5	302	21	AAY84320	Human cardiovascul
34	441	8.5	302	21	AAY59143	Human serine/threo
35	441	8.5	302	22	AAB71956	Human TGF-beta rec
36	441	8.5	311	22	AAG75475	Human colon cancer
37	439	8.5	302	22	AAB50054	Murine Dendritic C
38	436.5	8.4	369	21	AAY59145	Human serine/threo
39	432	8.3	268	21	AAY59146	Human serine/threo
40	432	8.3	306	23	AAU97064	Serine/threonine p
41	432	8.3	313	23	ABP61443	Human NF-kB activa
42	432	8.3	330	22	AAU87410	Novel central nerv
43	429.5	8.3	645	22	AAU03546	Human protein kina
44	429.5	8.3	645	23	AAE19147	Human kinase polyp
45	428.5	8.3	446	22	AAU03547	Human protein kina

ALIGNMENTS

RESULT 1
AAW40309
ID AAW40309 standard; Protein; 979 AA.
XX
AC AAW40309;
XX
DT 23-JUN-1998 (first entry)
XX
DE Human ITAK protein.
XX
KW Interleukin-1/tumour necrosis factor alpha activated kinase; ITAK; human;
KW inhibitor; cytokine mediated inflammation; antagonist; disorder; therapy;
KW IL-1; TNF.
XX
OS Homo sapiens.
XX
PN WO9747750-A1.
XX
PD 18-DEC-1997.
XX
PF 09-JUN-1997; 97WO-US08516.
XX
PR 10-JUN-1996; 96US-0633414.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Anderson DM, Bird TA, Sims JE, Virca G;
XX
DR WPI; 1998-052314/05.
DR N-PSDB; AAV10514; AAV10525.
XX
PT Nucleic acid encoding interleukin-1/tumour necrosis factor activated
PT kinase - used to identify specific antagonists for treatment of

```
PT   cytokine-mediated inflammation
XX
PS   Claim 3; Fig 1; 80pp; English.
XX
CC   This sequence represents an interleukin-1/tumour necrosis factor alpha
CC   activated kinase (ITAK). Antagonists of ITAK are used to treat IL-1 or
CC   TNF alpha-mediated inflammatory disorders e.g. rheumatoid arthritis,
CC   inflammatory bowel disease, type I diabetes, psoriasis, Alzheimer's
CC   disease, reperfusion injury, malignancy, transplant rejection, neuropathy
CC   associated with human immunodeficiency virus etc. Cells containing ITAK
CC   can be used to raise antibodies for assay of ITAK or to inhibit IL-1 and
CC   TNF alpha activity. Gene products that associate with ITAK are potential
CC   inhibitors and can be used to detect ITAK genes. Antisense sequences
CC   inhibit expression of ITAK. Inhibition of ITAK selectively blocks cell
CC   responses to IL-1 and TNF alpha, but not responses to other cytokines.
XX
SQ   Sequence   979 AA;

  Query Match            100.0%;  Score 5183;  DB 19;  Length 979;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  979;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979
              |||||||||||||||||||
Db        961 DLDSDSWCLLGTDSCRPSL 979


RESULT 2
AAY95294
ID   AAY95294 standard; Protein; 979 AA.
XX
AC   AAY95294;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Human GEF containing NEK-like kinase (GNK).
XX
KW   Human; GNK; GEF containing NEK-like kinase; GNK;
KW   IL-1/TNF-alpha activated kinase; ITAK; vascularization;
KW   vasculogenesis; blood vessel; angiogenesis; inflammation;
KW   arthritis; psoriasis; diabetic retinopathy; antiarthritic;
KW   antipsoriatic; cardiant; antiinflammatory; antidiabetic;
KW   opthalmological; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..43
FT                    /note= "leader peptide"
FT   Domain           44..315
FT                    /note= "kinase domain"
FT   Region           318..605
FT                    /note= "guanine nucleotide exchange factor homology
FT                          region"
FT   Region           752..764
FT                    /note= "glycine/acid-rich tether region"
FT   Domain           765..979
FT                    /note= "C-terminal cytoplasmic domain"
XX
PN   WO200036097-A2.
XX
PD   22-JUN-2000.
XX
PF   17-DEC-1999;   99WO-US29989.
XX
PR   18-DEC-1998;   98US-0113003.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Bird TA,  Peschon JJ,  Sims JE,  Virca CD,  Willis CR;
XX
DR   WPI; 2000-442384/38.
DR   N-PSDB; AAA27897.
XX
PT   Substrate for GEF-containing NEK-like Kinase (sGNK) nucleic acids,
PT   encoded proteins and antibodies, useful for modulation of
PT   vascularization and treatment of disorders such as arthritis, diabetic
PT   retinopathy, inflammation, and psoriasis -
XX
PS   Claim 10; Fig 4; 100pp; English.
XX
CC   The present sequence is that of human GEF-containing NEK-like
```

```
CC   kinase (GNK), a novel protein kinase, previously designated
CC   IL-1/TNF-alpha activated kinase (ITAK), that is involved in
CC   vascular development.  The invention is directed toward the use of
CC   GNK, and its physiological substrate sGNK (see AAY95293), in
CC   regulating vascularization.  It is especially directed to
CC   stimulating blood vessel development using GNK and its agonists,
CC   and to inhibiting inappropriate blood vessel development using
CC   antagonists of GNK, and hence for the treatment of disorders
CC   such as arthritis, diabetic retinopathy, inflammation and
CC   psoriasis.  GNK/sGNK polypeptides can also be used as purification
CC   reagents, to measure iological activity, to identify agonists and
CC   antagonists of GNK/sGNK; to identify unknown proteins; and to
CC   raise antibodies.
XX
SQ   Sequence   979 AA;

  Query Match             99.9%;  Score 5180;  DB 21;  Length 979;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  978;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEEEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979
              |||||||||||||||||||
Db        961 DLDSDSWCLLGTDSCRPSL 979


RESULT 3
AAB65703
ID   AAB65703 standard; Protein; 836 AA.
XX
AC   AAB65703;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Novel protein kinase, SEQ ID NO: 232.
XX
KW   Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW   immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW   dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW   immune disorder; cardiovascular disease; neurodegenerative disease;
KW   cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW   inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
OS   Homo sapiens.
XX
PN   WO200073469-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14842.
XX
PR   28-MAY-1999;   99US-0136503.
XX
PA   (SUGE-) SUGEN INC.
XX
PI   Plowman GD,  Martinez R,  Whyte D,  Sudersanam S;
XX
DR   WPI; 2001-032161/04.
DR   N-PSDB; AAF44732.
XX
PT   Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT   treating immune-related diseases and disorders, cardiovascular disease,
PT   neurodegenerative diseases and/or cancers -
XX
PS   Claim 10; Fig 1; 310pp; English.
XX
CC   The present sequence is a novel protein kinase. The novel protein kinases
CC   and the nucleic acids that encode them may be used in the treatment and
CC   diagnosis of diseases associated with inappropriate kinase expression
CC   such as immune-related diseases and disorders, cardiovascular disease,
CC   neurodegenerative diseases and/or cancers. The nucleic acids and
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays. The kinase polypeptides may be used as antigens in the production
CC   of antibodies of kinase expression and activity. Anti-kinase antibodies
CC   and kinase antagonists may also be used to down regulate kinase
CC   expression and activity. Diseases related to kinase expression and
CC   activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC   disorders, complications of organ transplantation, myocardial inf
CC   immune disorders, cardiomyopathies, strokes, renal failure,
CC   oxidative-stress related disorders, chronic inflammatory bow
CC   chronic inflammatory pelvic disease, multiple sclerosis, a
CC   osteoarthritis, psoriasis, rhinitis, autoimmunity, diabe
CC   reproductive disorders.
```

```
XX
SQ   Sequence   836 AA;

  Query Match           79.5%;  Score 4122;  DB 22;  Length 836;
  Best Local Similarity   88.6%;  Pred. No. 0;
  Matches  802;  Conservative    3;  Mismatches   16;  Indels   84;  Gaps    4;

Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||                ||||||||||||||||||||||||||||
Db        301 ELLDRPLLRKRR----------------RSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 343

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 403

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 EKLQGKAIRQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 463

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 523

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 583

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 643

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              |||||||
Db        644 FTIAATD---------------------------------------------------- 650

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 ----EKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 706

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        707 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 766

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRL----NPAVTCAGKGTPL------TPPAC 890
              ||||||||||||||||||||||||||||||    |    :  || |   : |      | :|
Db        767 DTLPYEELQGLKVASEAPLEHKPQVEASVTELFAFESQLVTSAESCSNLCWEGNTTDSSC 826

Qy        891 ACSSL 895
                |  |
Db        827 VCVQL 831


RESULT 4
```

```
AAW69341
ID   AAW69341 standard; Protein; 522 AA.
XX
AC   AAW69341;
XX
DT   25-NOV-1998  (first entry)
XX
DE   Secreted protein of clone CG279_1.
XX
KW   Secreted protein; nutritional activity; immune sti.
KW   immune suppressor; haematopoiesis regulator; tissue
KW   activin/inhibin activity; chemotactic/chemokinetic
KW   haemostatic activity; thrombolytic activity; recept                    ity;
KW   anti-inflammatory activity; cadherin suppressor; tum                   itor;
KW   tumour invasion suppressor; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9837094-A2.
XX
PD   27-AUG-1998.
XX
PF   24-FEB-1998;   98WO-US03595.
XX
PR   23-FEB-1998;   98US-0028168.
PR   24-FEB-1997;   97US-0804561.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-467491/40.
DR   N-PSDB; AAV58357.
XX
PT   Secreted proteins with biological activity - and encoding
PT   poly:nucleotide(s), useful e.g. in therapy and diagnosis of medical
PT   conditions and to identify protein agonists or antagonists
XX
PS   Claim 18; Page 73-74; 111pp; English.
XX
CC   This sequence is a secreted protein of the invention. This sequence
CC   is encoded by the DNA of clone CG279_1, which was isolated from a human
CC   adult testes cDNA library. The polypeptides are predicted to have
CC   useful biological activities which would make them suitable for treating,
CC   preventing or ameliorating medical conditions in humans and animals,
CC   although no supporting biological data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity or other activities. They may also be used for diagnostic
CC   purposes. The polynucleotides are useful to produce the proteins or
CC   generate probes or primers to identify and/or amplify similar genes
CC   e.g. species homologues. They are also useful for gene therapy and to
CC   produce transgenic animals with altered gene expression.
XX
SQ   Sequence   522 AA;

  Query Match           48.0%;  Score 2489;  DB 19;  Length 522;
  Best Local Similarity   99.6%;  Pred. No. 7.4e-203;
  Matches  468;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        510 GRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQC 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 GRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQC 69

Qy        570 MSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQL 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 MSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQL 129

Qy        630 GVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTER 689
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTER 189

Qy        690 PHGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLSIGTVFQSS 749
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 PHGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLSIGTVFQSS 249

Qy        750 SPGGGGGGGGGEEEDSQQESETPDPSGGFRGTMEADRGMEGLISPTEAMGNSNGASSSCP 809
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 SPGGGGGGGGGEEEDSQQESETPDPSGGFRGTMEADRGMEGLISPTEAMGNSNGASSSCP 309

Qy        810 GWLRKELENAEFIPMPDSPSPLSAAFSESEKDTLPYEELQGLKVASEAPLEHKPQVEASS 869
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 GWLRKELENAEFIPMPDSPSPLSAAFSESEKDTLPYEELQGLKVASEAPLEHKPQVEASS 369

Qy        870 PRLNPAVTCAGKGTPLTPPACACSSLQVEVERLQGLVLKCLAEQQKLQQENLQIFTQLQK 929
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 PRLNPAVTCAGKGTPLTPPACACSSLQVEVERLQGLVLKCLAEQQKLQQENLQIFTQLQK 429

Qy        930 LNKKLEGGQQVGMHSKGTQTAKEEMEMDPKPDLDSDSWCLLGTDSCRPSL 979
                 |||||||||||||||||||||||||||||| |||||||||:||||||
Db        430 LNKKLEGGQQVGMHSKGTQTAKEEMEMDPKPDFDSDSWCLLGTNSCRPSL 479


RESULT 5
ABB11588
ID   ABB11588 standard; peptide; 261 AA.
XX
AC   ABB11588;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human secreted protein homologue, SEQ ID NO:1958.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   N-PSDB; ABA08832.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 20; Page 217; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention.
XX
SQ   Sequence   261 AA;

  Query Match             26.0%;  Score 1347;  DB 22;  Length 261;
  Best Local Similarity   97.7%;  Pred. No. 4.6e-106;
  Matches  251;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        502 YSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEF 561
                 :|: ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FSFLCGVSGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEF 60

Qy        562 NKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTF 621
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTF 120

Qy        622 GCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGR 681
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGR 180

Qy        682 LAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLS 741
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLS 240

Qy        742 IGTVFQSSSPGGGGGGG 758
                 |||||||||||||| ||
Db        241 IGTVFQSSSPGGGGEGG 257


RESULT 6
AAO15418
ID   AAO15418 standard; Protein; 692 AA.
XX
AC   AAO15418;
XX
DT   27-SEP-2002  (first entry)
XX
DE   Novel human kinase protein 1.
```

```
XX
KW   Human; single nucleotide polymorphism; SNP; kinase protein;
KW   screening; mental disorder; biological disorder; medical disorder.
XX
OS   Homo sapiens.
XX
PN   WO200242438-A2.
XX
PD   30-MAY-2002.
XX
PF   19-NOV-2001; 2001WO-US43825.
XX
PR   20-NOV-2000; 2000US-252011P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Friddle CJ,  Hilbun E,  Mathur B,  Turner CA;
XX
DR   WPI; 2002-566563/60.
DR   N-PSDB; AAL44015.
XX
PT   Novel human kinase polynucleotide useful in therapeutic, diagnostic and
PT   pharmacogenomic applications -
XX
PS   Claim 2; Page 38-39; 43pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two novel
CC   human kinase proteins. The human kinase DNA and protein sequences of the
CC   invention are useful in therapeutic, diagnostic and pharmacogenomic
CC   applications. The DNA and protein sequences of the invention are useful
CC   for screening for compounds that can be used as pharmaceutical agents in
CC   the therapeutic treatment of mental, biological or medical disorders. The
CC   present amino acid sequence represents a novel human kinase protein of
CC   the invention.
XX
SQ   Sequence   692 AA;

  Query Match             20.9%;  Score 1084;  DB 23;  Length 692;
  Best Local Similarity   36.7%;  Pred. No. 5.5e-83;
  Matches  260;  Conservative  121;  Mismatches  269;  Indels   58;  Gaps   17;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL-------------LRKRRREMEEKVTLLNAPTKRPRSSTVTE 336
Db        241 LEPAQRPPLSHIMAQPLCIRALLNLHTDVGSVRMRRAEKSVAPSNTGSRTTSVRCRGIPR 300

Qy        337 API-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW-- 393
Db        301 GPVRPAIPPPLSSVYAWGGGLGTPLRLPMLNT--EVVQVAAGRTQKAGVTRSGRLILWEA 358

Qy        394 --VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCVTDEGQL 447
Db        359 PPLGAGGGSLLPGAV------EQPQPQFISRFLEGQSGVTIKHVACGDFFTACLTDRGII 412

Qy        448 YAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCG 507
Db        413 MTFGSGSNGCLGHGSLT--DISQPTIVEALLGYEMVQVACGASHVLALSTERELFAWGRG 470

Qy        508 EYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLN 567
Db        471 DSGRLGLGTRESHSCPQQVPMPPGQEAQRVVCGIDSSMILTVPGQALACGSNRFNKLGLD 530

Qy        568 QCMSG--IINHEAYHEVPYTTSFTL--AKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGC 623
Db        531 HLSLGEEPVPHQ---QVEEALSFTLLGSAPLDQEPLLSIDLGTAHSAAVTASGDCYTFGS 587

Qy        624 NKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRL 682
Db        588 NQHGQLGTNTRRGSRAPCKVQG--LEGIKMAMVACGDAFTVAIGAESEVYSWGKGARGRL 645

Qy        683 AMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVEKV 727
Db        646 GR--------RDEDAGLPRPVQLDETHPYTVTSVSCCHGNTLLAVRSV 685


RESULT 7
AAE22160
ID   AAE22160 standard; Protein; 692 AA.
XX
AC   AAE22160;
XX
DT   25-JUL-2002  (first entry)
XX
DE   Human protein kinase 18480.
XX
KW   Human; protein kinase; 18480; enzyme; antisense therapy; angiogenesis;
KW   cellular proliferative disorder; differentiative; cardiovascular; skin;
KW   neural; prostatic; brain; skeletal muscular; protein-protein interaction;
KW   signal transduction; immune; vascular; reproductive; atherosclerosis;
KW   rheumatoid arthritis; hypertension; coronary artery disease; arrhythmia;
KW   diabetes; ischaemic heart disease; angina pectoris; varicose vein; ulcer;
KW   Wegeners granulomatosis; cancer; wound healing; multiple sclerosis; HIV;
KW   Crohn's disease; asthma; allergy; infection; hepatitis; kidney disease;
KW   glomerulonephritis; idiopathic thrombocytopenic purpura; tuberculosis;
KW   Alzheimer's disease; chromosome mapping; drug screening; tissue typing;
KW   gene therapy; SLE.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          2..120
FT                   /note= "Kinase cell mitosis serine/threonine-protein
FT                   cycle 2.7.1.-phosphorylation division nuclear G2-specific
FT                   domain"
FT   Domain          2..120
FT                   /note= "Kinase serine/threonine-protein Y39G8B.5 III
FT                   R107.4 chromosome ATP-binding transferase 2.7.1 domain"
FT   Domain          4..258
FT                   /note= "Protein kinase domain"
FT   Binding-site    10..18
FT                   /note= "ATP-binding region signature site"
FT   Region          124..136
FT                   /note= "Serine/theronine kinase active site signature"
FT   Domain          182..204
FT                   /note= "Kinase serine/threonine-protein Y39G8B.5 III
FT                   R107.4 chromosome ATP-binding transferase 2.7.1 domain"
FT   Domain          325..437
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
FT   Domain          396..479
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
FT   Domain          445..540
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
FT   Binding-site    469..471
FT                   /note= "Cell attachment site"
FT   Domain          500..597
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
FT   Domain          560..658
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
FT   Domain          598..601
FT                   /note= "cAMP-dependent protein kinase phosphorylation
FT                   site"
```

```
FT   Domain          618..690
FT                   /note= "Guanine-nucleotide releasing cell cycle domain"
XX
PN   WO200206330-A2.
XX
PD   24-JAN-2002.
XX
PF   18-JUL-2001; 2001WO-US22820.
XX
PR   18-JUL-2000; 2000US-219028P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers R,  Kapeller-Libermann R,  Rudolph-Owen L,  Tsai F;
XX
DR   WPI; 2002-351402/38.
DR   N-PSDB; AAD35237.
XX
PT   New isolated human nucleic acid molecules, designated 13237, 18480,
PT   2245 or 16228, encode novel protein kinase family members, useful for
PT   the treatment and diagnosis of e.g. immune, cardiovascular,
PT   reproductive and cancerous disorders -
XX
PS   Claim 4; Fig 6; 159pp; English.
XX
CC   The invention relates to human nucleic acid molecules, designated 13237,
CC   18480, 2245 or 16228, which encode protein kinase family members. The
CC   proteins, polynucleotides, antibodies and modulators are useful in the
CC   treatment of cellular proliferative or differentiative, neural, skin,
CC   cardiovascular, brain, prostatic, and skeletal muscular disorders. They
CC   are also useful for the treatment and diagnosis of 13237, 18480, 2245 or
CC   16228-mediated disorders e.g. protein-protein interaction disorders,
CC   signal transduction disorders, immune (diabetes, rheumatoid arthritis,
CC   SLE), reproductive, cardiovascular (hypertension, atherosclerosis,
CC   coronary artery disease, arrhythmias, ischaemic heart disease, angina
CC   pectoris), vascular disorders (Wegeners granulomatosis, varicose veins,
CC   wound healing) or cancerous disorders, multiple sclerosis, Crohn's
CC   disease, ulcers, asthma allergy, infection, hepatitis, kidney disease
CC   (glomerulonephritis), idiopathic thrombocytopenic purpura, tuberculosis,
CC   HIV, Alzheimer's; for the diagnosis of a predisposition to a disorder;
CC   for evaluating the efficacy of a therapeutic or prophylactic disorder;
CC   for chromosome mapping; as immunogens; for drug screening; for the
CC   detection of mutations in the gene; for tissue typing and in gene
CC   therapy. The present sequence is human protein kinase 18480 protein.
XX
SQ   Sequence   692 AA;

  Query Match             20.9%;  Score 1081;  DB 23;  Length 692;
  Best Local Similarity   36.7%;  Pred. No. 1e-82;
  Matches  260;  Conservative  120;  Mismatches  270;  Indels   58;  Gaps   17;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:||||||   |  |  |  ||: |:: : :::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  | | :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALTIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL--------------LRKRRREMEEKVTLLNAPTKRPRSSTVTE 336
               :| |||    :: :||              :| || |     :   : |   |   :
Db        241 LEPAQRPPLSHIMAQPLCIRALLNLHTDVGSVRMRRAEKSVAPSNTGSRTTSVRCRGIPR 300

Qy        337 API-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW-- 393
               |:   :    | || ||||  || :| :: :     || || |  | ||    | `|
Db        301 GPVRPAIPPPLSSVYAWGGGLGTPLRLPMLNT--EVVQVAAGRTQKAGVTRSGRLILWEA 358

Qy        394 --VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCVTDEGQL 447
                 :   ||: | | :          ||: : :      |  |  |:||| || |:|| | :
Db        359 PPLGAGGGSLLPGAV------EQPQPQFISRFLEGQSGVTIKHVACGDFFTACLTDRGII 412

Qy        448 YAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCG 507
               |||   ||:|   :   :: :|  :   |   : ||:|| :||: |:  :|:::|| |
Db        413 MTFGSGSNGCLGHGSLT--DISQPTIVEALLGYEMVQVACGASHVLALSTERELFAWGRG 470

Qy        508 EYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLN 567
              : ||||| : | :  ||:| :|       | || | : :||  |: |||| | ||||||:
Db        471 DSGRLGLGTRESHSCPQQVPMPPGQEAQRVVCGIDSSMILTVPGQALACGSNRFNKLGLD 530

Qy        568 QCMSG--IINHEAYHEVPYTTSFTL--AKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGC 623
                |   : |:   :|    |||| :  |    : :|  |  |:||:   |   |||
Db        531 HLSLGEEPVPHQ---QVEEALSFTLLGSAPLDQEPLLSIDLGTAHSAAVTASGDCYTFGS 587

Qy        624 NKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRL 682
              |: ||||    : |    : |  | | ::  |:||| ||:|   :: :::|| |  |||
Db        588 NQHGQLGTNTRRGSRAPCKVQG--LEGIKMAMVACGDAFTVAIGAESEVYSWGKGARGRL 645

Qy        683 AMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVEKV 727
                        |    |||:    |   |  :||   :|:| |  |
Db        646 GR--------RDEDAGLPRPVQLDETHPYTVTSVSCCHGNTLLAVRSV 685


RESULT 8
ABB65338
ID   ABB65338 standard; Protein; 841 AA.
XX
AC   ABB65338;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 22806.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL09441.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 22806; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   841 AA;

  Query Match             16.4%;  Score 848;  DB 22;  Length 841;
  Best Local Similarity   31.2%;  Pred. No. 9.6e-63;
  Matches  223;  Conservative  125;  Mismatches  267;  Indels  100;  Gaps   21;

Qy         38 PR-AGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDA 96
              ||  | | |  |  :|  :|:|:|:|| | |||| |   :|:|:::|: ||   |  |
Db         90 PRETGVGVAGPELANYEKVRVVGQGSFGIAILYRRKSDGHQIVFKQINLSELSPPGRDLA 149

Qy         97 LNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKL-FEEEMVVW 155
              :||: : : | | ||::|   |: : |||||:|| :|| |   |  :: || | |  ::
Db        150 MNEVDVFSKLHHPNIVSYLGSFIKDNTLLIEMEYADGGTLAQIIAERQGKLHFPERYIIA 209

Qy        156 YLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTP 215
                 || ||:: :|   |||||:|| |:|| :  ::|:||:|:::| :|::   |:|::|||
Db        210 VFEQISSAINYMHSENILHRDLKTANVFLNRRGIVKIGDFGISKIMNTKIH-AQTVLGTP 268

Qy        216 YYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSS 275
              || |||:|:| :|: |||||:||:: |:  ||:|| |:|   |  ||: |      | |
Db        269 YYFSPEMCEGKEYDNKSDIWALGCILGEMCCLKKTFAASNLSELVTKIMAG-NYTPVPSG 327

Qy        276 QYSLELIQMVHSCLDQDPEQRPTADE-------LLDRPLLRKRRREMEEKV-----TLLN 323
               |:  |  :: : |  :  :|||| |       |: | | : :    |: |       |
Db        328 -YTSGLRSLMSNLLQVEAPRRPTASEVLVYWIPLIFRSLGKNKGYSYEDDVGGPGSDQLT 386

Qy        324 APTKRPRSSTV--------------------TEAPIAVVTSRTSEVYVWGGGK------- 356
               ||      | |                   | ||  ::  |:    :   |
Db        387 APVPAAAYSNVSMELELPTAQTETKQLMIADTAAPHEILEKRSVLYQLKAFGTCFSMAPI 446

Qy        357 STPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQ 416
               | |  ::       |   ::|| ||  :   | |     |   |||||     :::
Db        447 QLPPKAVIV-------DVAMSDSHFVVVNEDGSAYAW-----GEGTHGQLGLTALEAWKH 494

Qy        417 -PKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLN 475
               |  :| ::   :     || ||: ||  | | : ||: :  :| |:        |  :
Db        495 YPSRMESVRNYHVVSACAGDGFTILVTQAGSLLSCGSNAHLALGQDEQRNYH--SPKLIA 552

Qy        476 FFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIV 535
                 ||||: |  ||: |:|   || ||    | |||    :|   ||   |: :::
Db        553 RLADVRVEQVAAGLQHVLALSREGAVYVWGTSTCGALGLG---NYQQQQK--FPQKILLS 607

Qy        536 AVQ-------CGCDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSF 588
               |:       || | : :|  :|::  || |::|||  :              :
Db        608 HVKTKPSKIYCGPDTSAVLFANGELHVCGSNDYNKLGFQR----------------SAKI 651

Qy        589 TLAKQLSF-YKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGP- 646
               |  |::   :|:       ||:  : | | : | | |  || |: :       | :  |
Db        652 TAFKKVQLPHKVTQACFSSTHSVFLVEGGYVYTMGRNAEGQRGIRH------CNSVDHPT 705

Qy        647 ----LGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICT 697
                  :  : ::: :| |: || |::|| |  ||   ||   :  |    |  |||
Db        706 LVDSVKSRYIVKANCSDQCTIVASEDNIITVWGT-RNGLPGIGSTNCGLGLQICT 759


RESULT 9
AAU01056
ID   AAU01056 standard; Protein; 146 AA.
XX
AC   AAU01056;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secreted protein sequence encoded by gene #17.
XX
KW   Human secreted protein; gene therapy; autoimmune disease;
KW   hyperproliferative disorder; cardiovascular disorder;
KW   cerebrovascular disorder; nervous system disorder; infection;
KW   ocular disorder; wound healing; epithelial cell proliferation;
KW   skin aging; transplantation; tissue regeneration; chemotaxis;
KW   food additive; preservative.
XX
OS   Homo sapiens.
XX
PN   WO200123402-A1.
XX
PD   05-APR-2001.
XX
PF   26-SEP-2000; 2000WO-US26376.
XX
PR   27-SEP-1999;   99US-0155808.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-266138/27.
DR   N-PSDB; AAS01476.
XX
PT   Nucleic acids encoding 43 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 11; Page 471-472; 516pp; English.
XX
CC   AAU01040-AAU01082 represent novel human secreted proteins. The invention
CC   relates to these 43 human secreted proteins and their gene sequences
CC   (AAS01460-AAS01502) which can be used in gene therapy. The secreted
CC   proteins are useful to prevent, treat or ameliorate a medical condition
CC   in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC   sheep. The secreted proteins are also useful in diagnosing a pathological
CC   condition or susceptibility to a pathological condition. Antibodies to
CC   the secreted proteins can also be used in alleviating symptoms associated
CC   with disorders and in diagnostic immunoassays e.g. radioimmunoassays or
CC   enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed
CC   or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Parkinson's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. The polypeptides can also be
CC   used to aid wound healing and epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities.
XX
SQ   Sequence   146 AA;

  Query Match             14.1%;  Score 731;  DB 22;  Length 146;
  Best Local Similarity   97.3%;  Pred. No. 5.4e-54;
  Matches  142;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        152 MVVWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETL 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVVWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETL 60

Qy        212 VGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME 120

Qy        272 VDSSQYSLELIQMVHSCLDQDPEQRP 297
              |||||||||||||||||||  |  |
Db        121 VDSSQYSLELIQMVHSCLDQVQEPLP 146


RESULT 10
AAU01090
ID   AAU01090 standard; Protein; 146 AA.
XX
AC   AAU01090;
XX
DT   18-JUL-2001  (first entry)
```

```
XX
DE   Gene 17 Human secreted protein homologous amino acid sequence.
XX
KW   Human secreted protein; gene therapy; autoimmune disease;
KW   hyperproliferative disorder; cardiovascular disorder;
KW   cerebrovascular disorder; nervous system disorder; infection;
KW   ocular disorder; wound healing; epithelial cell proliferation;
KW   skin aging; transplantation; tissue regeneration; chemotaxis;
KW   food additive; preservative.
XX
OS   Homo sapiens.
XX
PN   WO200123402-A1.
XX
PD   05-APR-2001.
XX
PF   26-SEP-2000; 2000WO-US26376.
XX
PR   27-SEP-1999;   99US-0155808.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-266138/27.
XX
PT   Nucleic acids encoding 43 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Disclosure; Page 491; 516pp; English.
XX
CC   AAU01083-AAU01105 represent human secreted protein homologous amino acid
CC   sequences. They are described in the invention of 43 novel human secreted
CC   proteins (AAU01040-AAU01082) and their gene sequences (AAS01460-AAS01502)
CC   which can be used in gene therapy. The secreted proteins are useful to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. The secreted
CC   proteins are also useful in diagnosing a pathological condition or
CC   susceptibility to a pathological condition. Antibodies to the secreted
CC   proteins can also be used in alleviating symptoms associated with
CC   disorders and in diagnostic immunoassays e.g. radioimmunoassays or
CC   enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed
CC   or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Parkinson's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. The polypeptides can also be
CC   used to aid wound healing and epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities.
XX
SQ   Sequence   146 AA;

  Query Match           13.9%;  Score 718;  DB 22;  Length 146;
  Best Local Similarity   97.2%;  Pred. No. 6.9e-53;
  Matches  139;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        155 WYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGT 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 WYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGT 60

Qy        215 PYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDS 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDS 120

Qy        275 SQYSLELIQMVHSCLDQDPEQRP 297
              |||||||||||||||  |  |
Db        121 SQYSLELIQMVHSCLDQVQEPLP 143

RESULT 11
AAM39211
ID   AAM39211 standard; Protein; 1214 AA.
XX
AC   AAM39211;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2356.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58367.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2356; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   1214 AA;

  Query Match           11.6%;  Score 599.5;  DB 22;  Length 1214;
  Best Local Similarity   37.5%;  Pred. No. 2.5e-41;
  Matches  124;  Conservative   71;  Mismatches  113;  Indels   23;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
```

```
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              ||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| :  || :||: |  :    |  || :|  :|     :
Db        184 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :        ||       |      | ||| |  :::: 
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 301

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::      |   | |   :|
Db        302 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 332
```

RESULT 12
AAU07102
ID AAU07102 standard; Protein; 1214 AA.
XX
AC AAU07102;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel human protein, NHP #2.
XX
KW Human; novel human protein; NHP; breast cancer; prostate cancer;
KW immunogen; antibody; gene therapy; antisense.
XX
OS Homo sapiens.
XX
PN WO200161016-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05356.
XX
PR 18-FEB-2000; 2000US-0183582.
PR 22-FEB-2000; 2000US-0184014.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Walke DW, Hu Y, Nepomnichy B, Turner CA, Zambrowicz B;
XX
DR WPI; 2001-502793/55.
DR N-PSDB; AAS11558.
XX
PT Isolated nucleic acids encoding novel human proteins useful for the
PT treatment of disease and as probes for testing and detection -
XX
PS Claim 4; Page 37-39; 69pp; English.
XX
CC The invention relates to novel human proteins (NHP) and the nucleic
CC acids encoding them. The nucleic acids encode mammalian transporter
CC proteins and are useful for the treatment (e.g. by gene therapy or
CC antisense technology) of any of a wide variety of symptoms associated
CC with biological disorders (e.g. breast and prostate cancer) or imbalances
CC and as probes for the identification, selection and validation of novel
CC molecular targets for drug discovery. The proteins may be used to raise
CC anti-NHP antibodies. The present sequence represents an NHP of the
CC invention.
XX
SQ Sequence 1214 AA;

Query Match 11.6%; Score 599.5; DB 22; Length 1214;
Best Local Similarity 37.5%; Pred. No. 2.5e-41;
Matches 124; Conservative 71; Mismatches 113; Indels 23; Gaps 5;

```
Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              ||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| :  || :||: |  :    |  || :|  :|     :
Db        184 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :        ||       |      | ||| |  :::: 
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 301

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::      |   | |   :|
Db        302 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 332
```

RESULT 13
AAM39210
ID AAM39210 standard; Protein; 1242 AA.
XX
AC AAM39210;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2355.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58366.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

```
XX
PS   Example 4; SEQ ID NO 2355; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   1242 AA;

  Query Match             11.6%;  Score 599.5;  DB 22;  Length 1242;
  Best Local Similarity   37.5%;  Pred. No. 2.6e-41;
  Matches  124;  Conservative   71;  Mismatches  113;  Indels   23;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              ||||||: |||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| :  || :||: |  :    |  || :|  :|     :
Db        184 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :         ||       |      | |||  |  ::::
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 301

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::      |   | |   :|
Db        302 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 332


RESULT 14
AAE21728
ID   AAE21728 standard; Protein; 460 AA.
XX
AC   AAE21728;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human PKIN-23 protein.
XX
KW   Human; kinase; enzyme; PKIN-23 protein; immune system disorder; anaemia;
KW   acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;
KW   asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
KW   AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
KW   Down's syndrome; gene therapy; protein therapy; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          4..258
FT                   /note= "Eukaryotic protein kinase domain"
FT   Domain          6..247
FT                   /note= "Protein kinase domain"
FT   Domain          7..247
FT                   /note= "Protein kinase domain"
FT   Domain          9..247
FT                   /note= "Protein kinase domain"
FT   Domain          10..120
FT                   /note= "Protein kinase domain"
FT   Domain          80..93
FT                   /note= "Tyrosine kinase catalytic domain"
FT   Domain          118..136
FT                   /note= "Tyrosine kinase catalytic domain"
FT   Domain          124..247
FT                   /note= "Protein kinase domain"
FT   Domain          184..206
FT                   /note= "Tyrosine kinase catalytic domain"
FT   Domain          227..249
FT                   /note= "Tyrosine kinase catalytic domain"
XX
PN   WO200218557-A2.
XX
PD   07-MAR-2002.
XX
PF   31-AUG-2001; 2001WO-US27219.
XX
PR   31-AUG-2000; 2000US-229873P.
PR   08-SEP-2000; 2000US-231357P.
PR   14-SEP-2000; 2000US-232654P.
PR   22-SEP-2000; 2000US-234902P.
PR   29-SEP-2000; 2000US-236499P.
PR   06-OCT-2000; 2000US-238389P.
PR   13-OCT-2000; 2000US-240542P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Bandman O,  Nguyen DB,  Walia NK,  Hafalia AJA,  Yao MG,  Gandhi AR;
PI   Gururajan R,  Ding L,  Patterson C,  Yue H,  Baughn MR,  Tribouley CM;
PI   Thornton M,  Elliott VS,  Lu Y,  Ison CH,  Au-Young J,  Tang YT;
PI   Azimzai Y,  Burrill JD,  Marcus GA,  Zingler KA,  Lu DAM,  Lal PG;
PI   Ramkumar J,  Warren BA,  Kearney L,  Policky JL,  Thangavelu K;
PI   Burford N;
XX
DR   WPI; 2002-329769/36.
DR   N-PSDB; AAD34320.
XX
PT   New human kinases, useful for diagnosing, treating or preventing immune
PT   system disorders (e.g. Crohn's disease), neurological disorders (e.g.
PT   epilepsy), or cell proliferative disorders (e.g. cancers such as
PT   leukemia or lymphoma)  -
XX
PS   Claim 78; Page 188-189; 218pp; English.
XX
CC   The present invention relates to human kinases (PKIN) and polynucleotides
CC   encoding such proteins. PKIN sequences of the invention are useful for
CC   diagnosing, treating or preventing disorders associated with aberrant
CC   expression of PKIN, particularly immune system disorders (e.g. acquired
CC   immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,
CC   anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-
CC   Tooth disease or seizures), cell proliferative disorders (e.g. cancers
CC   such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),
CC   and developmental disorders (e.g. Down's syndrome). They are also used
CC   in gene therapy and protein therapy. The present sequence is human
CC   PKIN-23 protein.
XX
SQ   Sequence   460 AA;

  Query Match             10.8%;  Score 558;  DB 23;  Length 460;
  Best Local Similarity   41.3%;  Pred. No. 1.8e-38;
  Matches  130;  Conservative   58;  Mismatches   97;  Indels   30;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:||||||   |  |  |  ||: |:: : :::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63
```

```
Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              ||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||||:|||::|| :||| |:| |  | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPT-------KRPRSSTVTEAPIAVVT 343
              :| |||    :: :||  :          |||  |       :||     :   ||:
Db        241 LEPAQRPPLSHIMAQPLCIR---------ALLNLHTDVGSVRMRRP-----VQGQRAVLG 286

Qy        344 SRTSEVYVWGGGKST 358
              |     ||    ||
Db        287 GR-----VWAPSGST 296


RESULT 15
AAB65702
ID   AAB65702 standard; Protein; 698 AA.
XX
AC   AAB65702;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Novel protein kinase, SEQ ID NO: 231.
XX
KW   Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW   immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW   dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW   immune disorder; cardiovascular disease; neurodegenerative disease;
KW   cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW   inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
OS   Homo sapiens.
XX
PN   WO200073469-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14842.
XX
PR   28-MAY-1999;   99US-0136503.
XX
PA   (SUGE-) SUGEN INC.
XX
PI   Plowman GD,  Martinez R,  Whyte D,  Sudersanam S;
XX
DR   WPI; 2001-032161/04.
DR   N-PSDB; AAF44731.
XX
PT   Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT   treating immune-related diseases and disorders, cardiovascular disease,
PT   neurodegenerative diseases and/or cancers -
XX
PS   Claim 10; Fig 1; 310pp; English.
XX
CC   The present sequence is a novel protein kinase. The novel protein kinases
CC   and the nucleic acids that encode them may be used in the treatment and
CC   diagnosis of diseases associated with inappropriate kinase expression
CC   such as immune-related diseases and disorders, cardiovascular disease,
CC   neurodegenerative diseases and/or cancers. The nucleic acids and
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays. The kinase polypeptides may be used as antigens in the production
CC   of antibodies of kinase expression and activity. Anti-kinase antibodies
CC   and kinase antagonists may also be used to down regulate kinase
CC   expression and activity. Diseases related to kinase expression and
CC   activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC   disorders, complications of organ transplantation, myocardial infarction,
CC   immune disorders, cardiomyopathies, strokes, renal failure,
CC   oxidative-stress related disorders, chronic inflammatory bowel disease,
CC   chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC   osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC   reproductive disorders.
XX
SQ   Sequence   698 AA;

  Query Match             10.7%;  Score 555;  DB 22;  Length 698;
  Best Local Similarity   31.2%;  Pred. No. 6.4e-38;
  Matches  152;  Conservative   73;  Mismatches  144;  Indels  118;  Gaps   12;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:|||||   |  |  |  ||: |:: : :::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              ||||:|| || | |  : :|:||:|::| |:|:        :|| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK--------STPCYISPELCEGKPYNQ 175

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        176 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 233

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVY 350
              :| |||    :: :||  :          |||  |
Db        234 LEPAQRPPLSHIMAQPLCIR---------ALLNLHT---------------------- 260

Qy        351 VWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKE----LYTWVNMQGGTKLHGQL 406
                                  |:|     |      |::    |   : |  |:  :| |
Db        261 ------------------DGREVRGPQQH-----REQDHQCPLQRGIIMTFGSGSNGCL 296

Qy        407 GHGDKASYRQPKHVEKLQG-KAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAG 465
              |||      ||  || | | : : ||     ||:                  |   :|:
Db        297 GHGSLTDISQPTIVEALLGYEMVQQVEEALSFTL--------------LGSAPLDQ--- 338

Qy        466 PEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRL----------GLD 515
                          |:  :  |  |   :|  :::   | |:  ||          |:
Db        339 --------------EPLLSIDLGTAHSAAVTGEEDL---GSGDVNRLPSWERGHLLAGVA 381

Qy        516 SEEDYYT 522
              |  |  |
Db        382 SSTDVST 388


Search completed: July  1, 2003, 08:35:11
Job time : 43.1808 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 33.8192 Seconds
(without alignments)
3246.633 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4095	100.0	824	21	AAY95293	Human GEF containi
2	4092	99.9	855	22	AAM78466	Human protein SEQ
3	4092	99.9	855	22	AAM38741	Human polypeptide
4	4092	99.9	870	22	ABG30208	Novel human diagno
5	4092	99.9	873	22	ABG17472	Novel human diagno
6	4092	99.9	873	22	AAM79450	Human protein SEQ
7	4092	99.9	873	22	AAM40527	Human polypeptide
8	4053	99.0	847	22	ABG17278	Novel human diagno
9	1330	32.5	782	22	ABB62515	Drosophila melanog
10	622.5	15.2	403	23	ABP41422	Human ovarian anti
11	420	10.3	170	20	AAW95350	Human foetal brain
12	362.5	8.9	2117	22	AAU32040	Novel human secret
13	362.5	8.9	2192	18	AAW21732	LexA/NuMA fusion p
14	362.5	8.9	2272	18	AAW21731	GAL4/HA/NuMA fusio
15	361	8.8	440	22	ABG30206	Novel human diagno
16	361	8.8	548	22	ABG17473	Novel human diagno
17	360.5	8.8	2115	21	AAY49937	Human NuMA protein
18	358	8.7	2779	22	ABB62371	Drosophila melanog
19	352.5	8.6	2207	22	AAU32041	Novel human secret
20	352	8.6	1857	23	AAU84350	Protein MYH11 diff
21	346	8.4	1879	22	AAM25750	Human protein sequ
22	346	8.4	1988	22	AAM40999	Human polypeptide
23	346	8.4	1988	22	AAM41000	Human polypeptide
24	341	8.3	2246	22	ABG05850	Novel human diagno
25	340	8.3	2056	22	ABB59344	Drosophila melanog
26	335.5	8.2	1972	17	AAW00024	Smooth muscle myos
27	335	8.2	1690	22	ABB61144	Drosophila melanog
28	335	8.2	1690	22	ABB61173	Drosophila melanog
29	335	8.2	1960	22	AAM78854	Human protein SEQ
30	335	8.2	2143	22	ABG01716	Novel human diagno
31	333	8.1	1489	22	ABB59948	Drosophila melanog
32	333	8.1	2017	22	ABG06301	Novel human diagno
33	332.5	8.1	2101	21	AAY49936	Human NuMA protein
34	331	8.1	2067	22	ABB71125	Drosophila melanog
35	330.5	8.1	1963	22	AAM79838	Human protein SEQ
36	329.5	8.0	2101	15	AAR47173	Sequence of the in
37	329.5	8.0	2101	22	AAG65799	Nuclear mitotic ap
38	326	8.0	2383	23	ABG65631	Human breast speci
39	326	8.0	2663	22	AAM39097	Human polypeptide
40	326	8.0	2688	22	AAM40883	Human polypeptide
41	325.5	7.9	1154	22	AAU32042	Novel human secret
42	325.5	7.9	2442	21	AAY77575	Human cytoskeletal
43	325	7.9	1851	22	ABG01723	Novel human diagno
44	325	7.9	2633	22	ABG06505	Novel human diagno
45	324.5	7.9	1975	22	ABB62094	Drosophila melanog

ALIGNMENTS

RESULT 1
AAY95293
ID AAY95293 standard; Protein; 824 AA.
XX
AC AAY95293;
XX
DT 12-SEP-2000 (first entry)
XX
DE Human GEF containing NEK-like kinase substrate sGNK.
XX
KW Human; sGNK; GEF containing NEK-like kinase; GNK substrate;
KW vascularization; vasculogenesis; blood vessel; angiogenesis;
KW inflammation; arthritis; psoriasis; diabetic retinopathy;
KW antiarthritic; antipsoriatic; cardiant; antiinflammatory;
KW antidiabetic; opthalmological; therapy.
XX
OS Homo sapiens.
XX
PN WO200036097-A2.
XX
PD 22-JUN-2000.
XX
PF 17-DEC-1999; 99WO-US29989.
XX
PR 18-DEC-1998; 98US-0113003.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Bird TA, Peschon JJ, Sims JE, Virca CD, Willis CR;
XX
DR WPI; 2000-442384/38.
DR N-PSDB; AAA27896.
XX

PT Substrate for GEF-containing NEK-like Kinase (sGNK) nucleic acids,
PT encoded proteins and antibodies, useful for modulation of
PT vascularization and treatment of disorders such as arthritis, diabetic
PT retinopathy, inflammation, and psoriasis -
XX
PS Claim 8; Fig 2; 100pp; English.
XX
CC The present sequence is that of the physiological substrate (sGNK)
CC of human GEF-containing NEK-like kinase (GNK), a novel protein
CC kinase that is involved in vascular development. sGNK copurifies
CC with GNK on gel filtration. The present sequence was deduced from
CC isolated cDNA clones (see AAA27896). The invention is directed
CC toward the use of GNK and sGNK in regulating vascularization. It
CC is especially directed to stimulating blood vessel development
CC using GNK and its agonists, and to inhibiting inappropriate blood
CC vessel development using antagonists of GNK, and hence for the
CC treatment of disorders such as arthritis, diabetic retinopathy,
CC inflammation and psoriasis. GNK/sGNK polypeptides can also be
CC used as purification reagents, to measure iological activity, to
CC identify agonists and antagonists of GNK/sGNK; to identify unknown
CC proteins; and to raise antibodies.
XX
SQ Sequence 824 AA;

Query Match 100.0%; Score 4095; DB 21; Length 824;
Best Local Similarity 100.0%; Pred. No. 3.2e-226;
Matches 824; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||||
Db        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
```

RESULT 2
AAM78466
ID AAM78466 standard; Protein; 855 AA.
XX
AC AAM78466;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1128.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51599.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3361-3362; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX

```
SQ   Sequence   855 AA;

  Query Match             99.9%;  Score 4092;  DB 22;  Length 855;
  Best Local Similarity   99.9%;  Pred. No. 4.9e-226;
  Matches  823;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 824


RESULT 3
AAM38741
ID   AAM38741 standard; Protein; 855 AA.
XX
AC   AAM38741;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1886.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI57897.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 3; SEQ ID NO 1886; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   855 AA;

  Query Match             99.9%;  Score 4092;  DB 22;  Length 855;
  Best Local Similarity   99.9%;  Pred. No. 4.9e-226;
  Matches  823;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
```

```
Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 824
```

RESULT 4
ABG30208
ID ABG30208 standard; Protein; 870 AA.
XX
AC ABG30208;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #30199.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS94395.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 60567; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 870 AA;

Query Match 99.9%; Score 4092; DB 22; Length 870;
Best Local Similarity 99.9%; Pred. No. 5e-226;
Matches 823; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 75

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 135

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 195

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 255

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 315

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 375

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 435

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 495

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 555
```

```
Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 615

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 675

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 735

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 795

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        796 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 839
```

RESULT 5
ABG17472
ID ABG17472 standard; Protein; 873 AA.
XX
AC ABG17472;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17463.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS81659.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 47831; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 873 AA;

Query Match 99.9%; Score 4092; DB 22; Length 873;
Best Local Similarity 99.9%; Pred. No. 5e-226;
Matches 823; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 78

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 138

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 198

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 258

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 318

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 378

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 438

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 498

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 558

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 618

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 678

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 738

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 798

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        799 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 842
```

RESULT 6

```
AAM79450
ID   AAM79450 standard; Protein; 873 AA.
XX
AC   AAM79450;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3096.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52583.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 251; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   873 AA;

  Query Match             99.9%;  Score 4092;  DB 22;  Length 873;
  Best Local Similarity   99.9%;  Pred. No. 5e-226;
  Matches  823;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 78

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 138

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 198

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 258

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 318

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 378

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 438

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 498

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 558

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 618

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 678

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 738

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 798

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        799 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 842


RESULT 7
AAM40527
ID   AAM40527 standard; Protein; 873 AA.
XX
AC   AAM40527;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5458.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
```

```
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C;  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI59683.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5458; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   873 AA;

  Query Match             99.9%;  Score 4092;  DB 22;  Length 873;
  Best Local Similarity   99.9%;  Pred. No. 5e-226;
  Matches  823;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 78

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 138

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 198

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 258

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 318

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 378

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 438

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 498

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 558

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 618

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 678

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 738

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 798

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||||||||||||:
Db        799 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 842


RESULT 8
ABG17278
ID   ABG17278 standard; Protein; 847 AA.
XX
AC   ABG17278;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #17269.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS81465.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 47637; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

```
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   847 AA;

  Query Match            99.0%;  Score 4053;  DB 22;  Length 847;
  Best Local Similarity  99.9%;  Pred. No. 8.3e-224;
  Matches  815;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          9 EYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDY 68
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDY 60

Qy         69 EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQL 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQL 120

Qy        129 RNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEEN 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEEN 180

Qy        189 ISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLK 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLK 240

Qy        249 TEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGL 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGL 300

Qy        309 AKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQD 368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQD 360

Qy        369 TQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVD 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVD 420

Qy        429 INGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLL 488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 INGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLL 480

Qy        489 EKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNET 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNET 540

Qy        549 PNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPS 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPS 600

Qy        609 PGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKD 668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKD 660

Qy        669 KEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMK 728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 KEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMK 720
```

```
Qy        729 LRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQ 788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQ 780

Qy        789 KLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||||:
Db        781 KLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 816


RESULT 9
ABB62515
ID   ABB62515 standard; Protein; 782 AA.
XX
AC   ABB62515;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 14337.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06618.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 14337; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   782 AA;

  Query Match            32.5%;  Score 1330;  DB 22;  Length 782;
  Best Local Similarity  39.1%;  Pred. No. 4.7e-68;
  Matches  326;  Conservative  148;  Mismatches  247;  Indels  112;  Gaps   16;

Qy         22 LRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEA 81
              |: ||:||: || : :    |:|:||||::|||| |: : |||| |:  | |::  :||
Db         18 LQMEVERLTRELDQVSSASAQSAQYGLSLLEEKSALQQKCEELETLYDNTRHELDITQEA 77

Qy         82 FGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENER 141
                 :  |: |     |  :|::|: ||||::|     :: :|: ||||||: |   ::| :|
Db         78 LTKFQTSQKVTNKTGIEQEDALLNESAARETSLNLQIFDLENELKQLRHELERVRNERDR 137

Qy        142 LASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQN 201
```

```
Db        138 MLQENSDFGRDKSDSEADRLRLKSELKDLKFRETRMLSEYSELEEENISLQKQVSSLRSS 197

Qy        202 QVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKEL 261

Db        198 QVEFEGAKHEIRRLTEEVELLNQQVDELANLKKIAEKQMEEALETLQGEREAKYALKKEL 257

Qy        262 SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPK 321

Db        258 DGHLNRESMYHISNLAYSI--RSNMEDNASNNSDGEE-----ENLALKRLEADLSTELKS 310

Qy        322 KEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381

Db        311 PDGTK-------CDLFSEIHLNELKKLEKQLESMESEKTHLTANLREAQTSLDKSQNELQ 363

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441

Db        364 NFMSRLALLAAHVDALVQL------------KKQIDVKEQG------------------- 392

Qy        442 AVAEAG----ELREQLKALRS------THEAREAQHAEEKGRYEAEGQALTEKVSLL-EK 490

Db        393 --KEGGQKKDELEQQLRALISQYANWFTLSAKEIDGLKTDIAELQKGLNYTDATTTLRNE 450

Qy        491 ASRQDRELLARLEKELKKVSDV---------AGETQGSLSVAQDELVTFSEELANLYHHV 541

Db        451 VTNLKNKLLATEQKSLDLQSDVQTLTHISQNAGQSLGS---ARSTLVALSDDLAQLYHLV 507

Qy        542 CMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGG 601

Db        508 CTVNGETPTRVLLDHKTDDMS----FENDSLTAIQSQFKSDVFIAKPQIVEDLQGLADS- 562

Qy        602 TGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQEL 661

Db        563 ----------------------VEIKKYVDTVSDQIKYLKTAVEHTIDMNKHKIRS-EG 598

Qy        662 GPAVDK----DKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYEN 717

Db        599 GDALEKVNTEEMEELQEQIVKLKSLLSVKREQIGTLRNVLKSNKQTAEVALTNLKSKYEN 658

Qy        718 EKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKT 777

Db        659 EKIIVSDTMSKLRNELRLLKEDAATFSSLRAMFAARCEEYVTQVDDLNRQLEAAEEEKKT 718

Qy        778 LNSLLRMAIQQKLALTQRLELLELDHEQ--------TRRGRA-KAAPKTKPAT 821

Db        719 LNQLLRLAVQQKLALTQRLEEMEMDREMRHVRRPMPAQRGTSGKSSFSTRPSS 771
```

```
RESULT 10
ABP41422
ID   ABP41422 standard; Protein; 403 AA.
XX
AC   ABP41422;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HODEC78, SEQ ID NO:2554.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 12p11.2-11.1.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   N-PSDB; ABQ54499.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 11; SEQ ID No 2554; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disor
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleoti
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and th
CC   polypeptides may be used as food additives or to prepare anti
CC   useful in disease diagnosis, drug targeting and phenotyping
CC   sequence represents a human ovarian antigen of the inventi
CC   Note: The sequence data for this patent did not form part
CC   specification, but was obtained in electronic format di
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  403 AA;

  Query Match            15.2%;  Score 622.5;  DB 23
  Best Local Similarity  39.0%;  Pred. No. 6.4e-28;
  Matches  146;  Conservative   48;  Mismatches  11

Qy        311 LPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISE
Db         78 LPISFSLSPPASHPSPPPRNPLLPPSTGAMAP

Qy        371 KQLEHTRGSLSEQQEKVTRLTENLSALRR
Db        121 TEIE--------------RLTKEL---

Qy        431 GPEILACKYHVAVAEAGELREQLK
Db        154 GLEILECKYRVAVTEVIDLKA

Qy        491 ASRQDRELLARLEKELKK
Db        214 TTKESGEKMAHMEKEL

Qy        551 RVMLDYYREGQGG
```

```
Db        274 RVMLDYYRQSR----VTRSGSLKGPDDPRGLLSPRLARRGVSSPVETRTSSEPVAKESTE 329

Qy        610 GSSLPSP-------------------LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTT 649

Db        330 ASKEPSPTKTPTISXXITAPPSSXVLDTSDIRKEPMNIYNLNAIIRDQIKHLQKAVDRSL 389

Qy        650 ELSRQRIASQELGP 663

Db        390 QLXRQRAAARELAP 403
```

RESULT 11
AAW95350
ID AAW95350 standard; Protein; 170 AA.
XX
AC AAW95350;
XX
DT 26-APR-1999 (first entry)
XX
DE Human foetal brain secreted protein fn53_4.
XX
KW Secreted protein; human; brain; fn53_4.
XX
OS Homo sapiens.
XX
PN WO9856909-A2.
XX
PD 17-DEC-1998.
XX
PF 08-JUN-1998; 98WO-US11822.
XX
PR 05-JUN-1998; 98US-0092722.
PR 11-JUN-1997; 97US-0873218.
XX
PA (GEMY-) GENETICS INST INC.
XX
PI Agostino MJ, Fechtel K, Howes SH, Jacobs K, LaVallie ER;
PI McCoy JM, Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-080899/07.
DR N-PSDB; AAV99729.
XX
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult testes, foetal kidney, adult thyroid or
PT adult retina cDNA libraries
XX
PS Claim 28; Page 84; 113pp; English.
XX
CC This is the amino acid sequence of fn53_4, a novel human
CC secreted protein predicted from the nucleotide sequence of a
CC human foetal brain cDNA clone (see AAV99729). Database searches
CC indicate some sequence similarity to known sequences. The
CC invention provides cDNA clones (see AAV99721-33) from human adult
CC thyroid, adult retina, adult testis, foetal kidney and foetal
CC brain that encode novel secreted proteins (see AAW95344-53). The
CC polynucleotides and proteins are predicted to have activities
CC which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional,
CC cytokine, cell proliferation/differentiation, immune stimulating
CC (e.g. as vaccines) or immune suppressing, haematopoiesis regulating,
CC tissue growth, activin/inhibin, chemotactic/chemokinetic,
CC haemostatic, thrombolytic, receptor/ligand, antiinflammatory,
CC cadherin/tumour invasion suppressor, and tumour inhibition
CC activities.
XX
SQ Sequence 170 AA;

Query Match 10.3%; Score 420; DB 20; Length 170;
Best Local Similarity 49.7%; Pred. No. 8.8e-17;
Matches 84; Conservative 38; Mismatches 47; Indels 0; Gaps 0;

```
Qy        353 MQ
              |
Db          1 M

Qy        413 !

Db         61

Qy        473

Db         12
```

RESULT 12
AAU32040
ID AAU32
XX
AC AAU3
XX
DT 18-D
XX
DE Nov
XX
KW Hum
KW st
KW in
XX
OS H
XX
PN
XX
PD -
XX
PF 16-APR-2001; 2
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 547; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 2117 AA;

Query Match 8.9%; Score 362.5; DB 22; Length 2117;
Best Local Similarity 22.3%; Pred. No. 3.6e-12;
Matches 204; Conservative 163; Mismatches 336; Indels 211; Gaps 36;

```
Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREK----------IQAAEYGLAVLEEK 54
              :::| |    :  | | | : :|:   :| |   ::             |||   | | |:
Db        540 QQQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEKQEATRQDHAQQLATAAEEREASLRER 599

Qy         55 HQLKLQFEELEVD----YEAIRSEMEQLKEAFGQAHTN-----HKKVAADGESREESLIQ 105
               | | || :     | :: :::   ||   | |:       :|    :  |
Db        600 DAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREKAELSRKVEELQACV 659

Qy        106 ESASKEQYYVR-KVLELQTELKQLRNVLTNTQSENERLASVAQELKE----INQNVEIQR 160
              |:| :||:  : :| ||: :|:  :   |    | ||:|    :|:|    : ::::: :
Db        660 ETARQEQHEAQAQVAELELQLRSEQQKAT----EKERVAQEKDQLQEQLQALKESLKVTK 715

Qy        161 GRLRDDIKEYKFREARLLQD----YSELEEENISL----QKQVSVLRQNQVEFEGLKHEI 212
              | |    :| | | |  |::     |||: |  ||    :::   | : :   :||:  :
Db        716 GSL----EEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARL 771

Qy        213 KRLEE----ETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSIN 268
              ::| |    ||| |  :| :|:  :  :| : |: :: :   ||:    ::| : |
Db        772 QQLGEAHQAETEVLRRELAEAMAAQHTAESECEQLVKEVAAWRERYEDSQQEEAQY---- 827

Qy        269 DSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPP 328
                    |  | :       : |            |   : : :  | |:
Db        828 ------------GAMFQEQLMTLKEECE-----------KARQELQEAKEKVAGI--- 859

Qy        329 SPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVT 388
                  :  ||| ||  |    :| ::    |  |  :|: : : :     ||  |||:
Db        860 ------ESHSELQISRQQ---NELAELHANLARALQQVQEKEVRAQKLADDLSTLQEKMA 910

Qy        389 RLTENLSALRRL--QASKERQTALDNEKDRDSHEDGDYYEVDINGPEILA-------CKY 439
              :: :: |  |  :| ::::||   |  ::    ||        || |     |
Db        911 ATSKEVARLETLVRKAGEQQETA-SRELVKEPARAGD------RQPEWLEEQQGRQFCST 963

Qy        440 HVAV----AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS------LLE 489
                |:     || ::  :|: ||:     : |  ||:|: | |   ||::          ||
Db        964 QAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVARLTQERGRAQADLALE 1023

Qy        490 KASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETP 549
              ||:| : |:  ||:  | :         | :|: |  |     :||| |          |
Db       1024 KAARAELEM--RLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLR------GLEAA 1075

Qy        550 NRVMLDYYRE-----------------GQG----GAGRTSPGGRTSPEARGRRSPI--- 584
                  |:  |:                    | |     |||| |   | |:    |: :
Db       1076 QIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEP---TGPKLEALRAEVSKL 1132

Qy        585 ------------LLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIA 632
                         | : | |  | ||:        |   :|   | :  :|
Db       1133 EQQCQKQQEQADSLERSLEAERASRAE------RDSALETLQGQLEEKAQE--------- 1177

Qy        633 IIRDQIKHLQAAV-DRTTELSRQRIASQELGPAVDKDKEAL---MEEILKLKSLLSTKRE 688
                  : | |:|:     ||:  |   |:   | |: |  :    :|  :  ||:|:  |
Db       1178 -----LGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLISSLEE 1232

Qy        689 QITTL-RTVLKANKQTAEVALANLKSKYENEKAMV---TETMMKLRNELKALKEDAATFS 744
              ::: | | ||:   :: |:            | :|   :|   ||   |: |: : |: |
Db       1233 EVSILNRQVLEKEGESKEL-----------KRLVMAESEKSQKLEERLRLLQAETASNS 1280

Qy        745 SLRAMFATRCDEYITQL-DEMQRQLAAAEDEKKTLNSLLRMA--IQQKLAL--------T 793
              :  |  ::   | :  | :| ::|  |:|: :: | |    |  : |:|
Db       1281 ARAAERSSALREEVQSLREEAEKQRVASENLRQELTSQAERAEELGQELKAWQEKFFQKE 1340

Qy        794 QRLELLELDHEQTR 807
              | |  |:|:|  |:
Db       1341 QALSTLQLEHTSTQ 1354
```

RESULT 13
AAW21732
ID AAW21732 standard; Protein; 2192 AA.
XX
AC AAW21732;
XX
DT 01-OCT-1997 (first entry)
XX
DE LexA/NuMA fusion protein.
XX
KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;
KW cell division; proliferation; antibody; Ab; detection;
KW malignant cell growth.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..87
FT /label= LexA_DNA_binding_domain
FT Peptide 88..94
FT /label= Polylinker
FT Protein 95..2192
FT /label= Residues_18-2116_of_NuMA
FT Region 285..1784
FT /label= Coiled_coil_region
XX
PN WO9640917-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09504.
XX
PR 07-JUN-1995; 95US-0478408.
XX
PA (UYYA) UNIV YALE.
XX
PI McPherson SMG, Snyder MP;
XX
DR WPI; 1997-077270/07.
DR N-PSDB; AAT77783.
XX
PT New nucleic acid encoding nuclear mitotic appts. interacting
PT proteins - useful for modulating cell division and proliferation and
PT in diagnosis
XX
PS Claim 15; Page 42-50; 78pp; English.
XX
CC The sequences given in AAW21731-32 represent fusion proteins which
CC contain NuMA (nuclear mitotic apparatus). The fusion proteins were used
CC in the identification of NuMA interecting proteins (NIP's) (see also
CC AAW21729-30). Compounds which interfere with the interaction of NuMA
CC with a known NIP are used to modulate cell division and/or
CC proliferation. Ab, raised conventionally using NIP-1 or -2 as immunogen,
CC are used to detect NIP (or their complexes) and to block their activity
CC for diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP
CC which may be markers for aberrant (including malignant) cell growth
CC (which can also be detected by nucleic acid sequencing). Also where
CC malignancy is related to defects in NuMA or NIP, it can be treated by
CC administration of the appropriate functional protein.
XX
SQ Sequence 2192 AA;

Query Match 8.9%; Score 362.5; DB 18; Length 2192;
Best Local Similarity 22.3%; Pred. No. 3.7e-12;
Matches 204; Conservative 163; Mismatches 336; Indels 211; Gaps 36;

```
Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREK----------IQAAEYGLAVLEEK 54
              :::| |    :  | | | : :|:   :| |   ::             |||   | | |:
Db        617 QQQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEKQEATRQDHAQQLATAAEEREASLRER 676

Qy         55 HQLKLQFEELEVD----YEAIRSEMEQLKEAFGQAHTN-----HKKVAADGESREESLIQ 105
               | | || :     | :: :::   ||   | |:       :|    :  |
Db        677 DAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREKAELSRKVEELQACV 736

Qy        106 ESASKEQYYVR-KVLELQTELKQLRNVLTNTQSENERLASVAQELKE----INQNVEIQR 160
              |:| :||:  : :| ||: :|:  :   |    | ||:|    :|:|    : ::::: :
Db        737 ETARQEQHEAQAQVAELELQLRSEQQKAT----EKERVAQEKDQLQEQLQALKESLKVTK 792

Qy        161 GRLRDDIKEYKFREARLLQD----YSELEEENISL----QKQVSVLRQNQVEFEGLKHEI 212
              | |    :| | | |  |::     |||: |  ||    :::   | : :   :||:  :
Db        793 GSL----EEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARL 848
```

```
Qy        213 KRLEE----ETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSIN 268
              ::| |    ||| |  :| :|:  :  :| : |: :: :   ||:    ::| : |
Db        849 QQLGEAHQAETEVLRRELAEAMAAQHTAESECEQLVKEVAAWRERYEDSQQEEAQY---- 904

Qy        269 DSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPP 328
                      |  | :       : |            |   : : :  |  |:
Db        905 -------------GAMFQEQLMTLKEECE-----------KARQELQEAKEKVAGI--- 936

Qy        329 SPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVT 388
                  :  ||| ||  |    :| ::    |  |  :|: : : :     ||  |||:
Db        937 ------ESHSELQISRQQ---NELAELHANLARALQQVQEKEVRAQKLADDLSTLQEKMA 987

Qy        389 RLTENLSALRRL--QASKERQTALDNEKDRDSHEDGDYYEVDINGPEILA-------CKY 439
              :: :: |  |  :| ::::||   |  ::    ||        || |        |
Db        988 ATSKEVARLETLVRKAGEQQETA-SRELVKEPARAGD------RQPEWLEEQQGRQFCST 1040

Qy        440 HVAV----AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS------LLE 489
              |:     || ::  :|: ||:     : |  ||:|: | |   ||::         ||
Db       1041 QAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVARLTQERGRAQADLALE 1100

Qy        490 KASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETP 549
              ||:| : |:  ||:  | :        | :|: |  |     :||| |        |
Db       1101 KAARAELEM--RLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLR------GLEAA 1152

Qy        550 NRVMLDYYRE-----------------GQG----GAGRTSPGGRTSPEARGRRSPI--- 584
              |:  |:                    | |     ||| |   | |:    |: :
Db       1153 QIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEP---TGPKLEALRAEVSKL 1209

Qy        585 ------------LLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIA 632
                         | : | |  | ||:        |   :|   | :  :|
Db       1210 EQQCQKQQEQADSLERSLEAERASRAE------RDSALETLQGQLEEKAQE--------- 1254

Qy        633 IIRDQIKHLQAAV-DRTTELSRQRIASQELGPAVDKDKEAL---MEEILKLKSLLSTKRE 688
                  : | |:|:     ||:  |   |:   | |: | :    :|  :  ||:|:  |
Db       1255 -----LGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLISSLEE 1309

Qy        689 QITTL-RTVLKANKQTAEVALANLKSKYENEKAMV---TETMMKLRNELKALKEDAATFS 744
              ::: | | ||:   :: |:             | :|   :|   ||   |: |: : |: |
Db       1310 EVSILNRQVLEKEGESKEL------------KRLVMAESEKSQKLEERLRLLQAETASNS 1357

Qy        745 SLRAMFATRCDEYITQL-DEMQRQLAAAEDEKKTLNSLLRMA--IQQKLAL--------T 793
              :  |  ::   | :  | :| ::|  |:|: :: | |    |  : |:|
Db       1358 ARAAERSSALREEVQSLREEAEKQRVASENLRQELTSQAERAEELGQELKAWQEKFFQKE 1417

Qy        794 QRLELLELDHEQTR 807
              | |  |:|:|  |:
Db       1418 QALSTLQLEHTSTQ 1431
```

```
RESULT 14
AAW21731
ID   AAW21731 standard; Protein; 2272 AA.
XX
AC   AAW21731;
XX
DT   01-OCT-1997  (first entry)
XX
DE   GAL4/HA/NuMA fusion protein.
XX
KW   NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;
KW   cell division; proliferation; antibody; Ab; detection;
KW   malignant cell growth.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..147
FT                   /label= GAL4_DNA_binding_domain
FT   Peptide         148..174
FT                   /label= Hemaglutinin_epitope
FT   Protein         175..2272
FT                   /label= Residues_18-2116_of_NuMA
FT   Region          365..1864
FT                   /label= Coiled_coil_region
XX
PN   WO9640917-A1.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09504.
XX
PR   07-JUN-1995;   95US-0478408.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   McPherson SMG,  Snyder MP;
XX
DR   WPI; 1997-077270/07.
DR   N-PSDB; AAT77782.
XX
PT   New nucleic acid encoding nuclear mitotic appts. interacting
PT   proteins - useful for modulating cell division and proliferation and
PT   in diagnosis
XX
PS   Claim 14; Page 28-36; 78pp; English.
XX
CC   The sequences given in AAW21731-32 represent fusion proteins which
CC   contain NuMA (nuclear mitotic apparatus).  The fusion proteins were used
CC   in the identification of NuMA interecting proteins (NIP's) (see also
CC   AAW21729-30).  Compounds which interfere with the interaction of NuMA
CC   with a known NIP are used to modulate cell division and/or
CC   proliferation. Ab, raised conventionally using NIP-1 or -2 as immunogen,
CC   are used to detect NIP (or their complexes) and to block their activity
CC   for diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP
CC   which may be markers for aberrant (including malignant) cell growth
CC   (which can also be detected by nucleic acid sequencing). Also where
CC   malignancy is related to defects in NuMA or NIP, it can be treated by
CC   administration of the appropriate functional protein.
XX
SQ   Sequence   2272 AA;
```

Query Match 8.9%; Score 362.5; DB 18; Length 2272;
Best Local Similarity 22.3%; Pred. No. 3.9e-12;
Matches 204; Conservative 163; Mismatches 336; Indels 211; Gaps 36;

```
Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREK-----------IQAAEYGLAVLEEK 54
              :::| |   :  | | | : :|:   :| |   ::             |||   | | |:
Db        697 QQQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEKQEATRQDHAQQLATAAEEREASLRER 756

Qy         55 HQLKLQFEELEVD----YEAIRSEMEQLKEAFGQAHTN-----HKKVAADGESREESLIQ 105
                 | | || :    | :: :::   ||   | |:      :|     :  |
Db        757 DAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREKAELSRKVEELQACV 816

Qy        106 ESASKEQYYVR-KVLELQTELKQLRNVLTNTQSENERLASVAQELKE----INQNVEIQR 160
              |:| :||:  : :| ||: :|:  :   |    | ||:|    :|:|    : ::::: :
Db        817 ETARQEQHEAQAQVAELELQLRSEQQKAT----EKERVAQEKDQLQEQLQALKESLKVTK 872

Qy        161 GRLRDDIKEYKFREARLLQD----YSELEEENISL----QKQVSVLRQNQVEFEGLKHEI 212
              | |    :| | | |  |::     |||: |  ||    :::   | : :   :||:  :
Db        873 GSL----EEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARL 928

Qy        213 KRLEE----ETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSIN 268
              ::| |   |||  |  :| :|:  :  :| : |: :: :   ||:    ::| : |
Db        929 QQLGEAHQAETEVLRRELAEAMAAQHTAESECEQLVKEVAAWRERYEDSQQEEAQY---- 984

Qy        269 DSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPP 328
                      |  | :       : |            |   : : :  |  |:
Db        985 -------------GAMFQEQLMTLKEECE-----------KARQELQEAKEKVAGI--- 1016

Qy        329 SPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVT 388
                  :  ||| ||  |    :| ::    |  |  :|: : : :     ||  |||:
Db       1017 ------ESHSELQISRQQ---NELAELHANLARALQQVQEKEVRAQKLADDLSTLQEKMA 1067

Qy        389 RLTENLSALRRL--QASKERQTALDNEKDRDSHEDGDYYEVDINGPEILA-------CKY 439
              :: :: |  |  :| ::::||   |  ::    ||        || |        |
```

```
Db       1068 ATSKEVARLETLVRKAGEQQETA-SRELVKEPARAGD------RQPEWLEEQQGRQFCST 1120

Qy        440 HVAV----AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS------LLE 489
                :     || ::  :|: ||:     : |  ||:|: | |   ||::          ||
Db       1121 QAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVARLTQERGRAQADLALE 1180

Qy        490 KASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETP 549
              ||:| : |:  ||:  | :         | :|: | |     :||| |          |
Db       1181 KAARAELEM--RLQNALNEQRVEFATLQEALAHALTEKEGKDQELAKLR------GLEAA 1232

Qy        550 NRVMLDYYRE------------------GQG----GAGRTSPGGRTSPEARGRRSPI--- 584
                |:  |:                   | |     |||| |   | |:    |: :
Db       1233 QIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEP---TGPKLEALRAEVSKL 1289

Qy        585 ------------LLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIA 632
                         | : | |  | ||:        |   :|   | :  :|
Db       1290 EQQCQKQQEQADSLERSLEAERASRAE------RDSALETLQGQLEEKAQE--------- 1334

Qy        633 IIRDQIKHLQAAV-DRTTELSRQRIASQELGPAVDKDKEAL---MEEILKLKSLLSTKRE 688
                  : | |:|:     ||:  |   |:   | |: |  :    :|  :  ||:|:  |
Db       1335 -----LGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLISSLEE 1389

Qy        689 QITTL-RTVLKANKQTAEVALANLKSKYENEKAMV---TETMMKLRNELKALKEDAATFS 744
              ::: | | ||:   :: |:           | :|   :|   ||   |: |: : |: |
Db       1390 EVSILNRQVLEKEGESKEL-----------KRLVMAESEKSQKLEERLRLLQAETASNS 1437

Qy        745 SLRAMFATRCDEYITQL-DEMQRQLAAAEDEKKTLNSLLRMA--IQQKLAL--------T 793
              :  |  ::   | :  | :| ::|  |:|: :: | |    |  : |:|
Db       1438 ARAAERSSALREEVQSLREEAEKQRVASENLRQELTSQAERAEELGQELKAWQEKFFQKE 1497

Qy        794 QRLELLELDHEQTR 807
              | |  |:|:|  |:
Db       1498 QALSTLQLEHTSTQ 1511
```

```
RESULT 15
ABG30206
ID   ABG30206 standard; Protein; 440 AA.
XX
AC   ABG30206;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #30197.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS94393.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 60565; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   440 AA;
```

```
  Query Match             8.8%;  Score 361;  DB 22;  Length 440;
  Best Local Similarity   45.4%;  Pred. No. 6.5e-13;
  Matches   99;  Conservative   31;  Mismatches   42;  Indels   46;  Gaps   11;

Qy         79 KEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSE 138
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 RQAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSE 142

Qy        139 NERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLL------QDYSELEEEN--IS 190
              ||||||||||||||    :| : |:| :      | || ::      :| :  |::
Db        143 NERLASVAQELKE---KIETE-GKLPN-----TFYEASIVLLPKPDKDINRKEKDKPVFL 193

Qy        191 LQKQVSVLRQ---NQVE--------------FEGLKH-EIKRLEEETEYLNSQLEDAIRL 232
              : |   || :   ||::                || | | | :.|:     | ::: | :
Db        194 MNKDAKVLNKILANQIQKCVKRIMLYDQGEFIPGLGHGEHKMILEQ---FASDIKEVITI 250

Qy        233 KEISERQLEEEALETLKTEREQKNSLRKELSHYMSINDS 270
                |:       : |||  :   :   :: :::||  |
Db        251 --ITN------YDVLKTWSQLGGAPTGQICYHLSIKAS 280
```

```
Search completed: July  1, 2003, 08:35:08
Job time : 37.8192 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 36.5613 Seconds
(without alignments)
4643.787 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4095	100.0	824	4	Q8TD16	Q8td16 homo sapien
2	4053	99.0	847	4	O75181	O75181 homo sapien
3	3875	94.6	820	11	Q921C5	Q921c5 mus musculu
4	3696	90.3	775	4	Q96LH2	Q96lh2 homo sapien
5	2925	71.4	586	4	Q9BT84	Q9bt84 homo sapien
6	2536	61.9	975	4	O43892	O43892 homo sapien
7	2528	61.7	835	11	Q91YP7	Q91yp7 mus musculu
8	1603.5	39.2	545	11	O55206	O55206 mus musculu
9	1196	29.2	264	11	Q9DCL3	Q9dcl3 mus musculu
10	869	21.2	257	11	Q8R2J6	Q8r2j6 mus musculu
11	860	21.0	869	5	Q94174	Q94174 caenorhabdi
12	491.5	12.0	172	4	O43893	O43893 homo sapien
13	396	9.7	80	4	Q9H561	Q9h561 homo sapien
14	382	9.3	3259	4	Q14789	Q14789 homo sapien
15	362.5	8.9	1229	5	Q9NJ22	Q9nj22 aequipecten
16	362.5	8.9	1243	5	Q9NJ21	Q9nj21 aequipecten
17	362.5	8.9	1253	5	Q9NJ20	Q9nj20 aequipecten
18	362.5	8.9	1951	5	Q17042	Q17042 aequipecten
19	362.5	8.9	2115	4	Q14980	Q14980 homo sapien
20	358.5	8.8	1219	5	Q9NJ23	Q9nj23 aequipecten
21	358	8.7	2779	5	Q9W4N7	Q9w4n7 drosophila
22	355	8.7	2139	5	Q07569	Q07569 entamoeba h
23	354.5	8.7	1941	5	Q26079	Q26079 placopecten
24	352.5	8.6	1790	3	Q07380	Q07380 saccharomyc
25	351.5	8.6	1940	5	Q9U7E3	Q9u7e3 pecten maxi
26	351.5	8.6	3187	11	Q63714	Q63714 rattus norv
27	348	8.5	1690	5	O44929	O44929 drosophila
28	348	8.5	1950	5	Q26080	Q26080 placopecten
29	347.5	8.5	2007	13	Q02015	Q02015 gallus gall
30	343	8.4	2017	5	Q94992	Q94992 drosophila
31	343	8.4	2057	5	Q94987	Q94987 drosophila
32	341	8.3	1960	11	Q8VDD5	Q8vdd5 mus musculu
33	340	8.3	2056	5	Q9W0W8	Q9w0w8 drosophila
34	338.5	8.3	1972	11	Q8R384	Q8r384 mus musculu
35	338	8.3	1489	5	Q8T8Q5	Q8t8q5 drosophila
36	337.5	8.2	1344	5	Q9XYL5	Q9xyl5 schmidtea m
37	337.5	8.2	1708	5	Q9U0S6	Q9u0s6 mytilus gal
38	337.5	8.2	1927	5	Q25142	Q25142 halocynthia
39	337	8.2	2138	5	Q9XZE3	Q9xze3 amoeba prot
40	335.5	8.2	1743	5	O96063	O96063 dugesia jap
41	335	8.2	1690	5	Q9VJE5	Q9vje5 drosophila
42	333	8.1	1175	5	Q9TY21	Q9ty21 drosophila
43	333	8.1	1489	5	Q9W1R3	Q9w1r3 drosophila
44	333	8.1	1978	4	O15154	O15154 homo sapien
45	333	8.1	1979	4	O95949	O95949 homo sapien

ALIGNMENTS

RESULT 1
Q8TD16
ID Q8TD16 PRELIMINARY; PRT; 824 AA.
AC Q8TD16;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Coiled-coil protein BICD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland P.M., Milne A., Garka K., Johnson R.S., Willis C.R.,
RA Sims J.E., Rauch C.T., Bird T.A., Virca G.D.;
RT "Purification, cloning and characterization of a novel NIMA-related
RT kinase, Nek8, and its candidate substrate Bicd2.";
RL J. Biol. Chem. 0:0-0(2002).
DR EMBL; AY052562; AAL12246.1; -.
SQ SEQUENCE 824 AA; 93533 MW; 9C49138FF416378D CRC64;

Query Match 100.0%; Score 4095; DB 4; Length 824;
Best Local Similarity 100.0%; Pred. No. 5.4e-159;
Matches 824; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
```

```
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Db        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Db        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Db        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Db        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Db        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Db        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Db        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Db        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Db        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824

Db        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
```

RESULT 2
O75181
ID O75181 PRELIMINARY; PRT; 847 AA.
AC O75181;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE KIAA0699 protein (Fragment).
GN KIAA0699.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014599; BAA31674.1; -.
FT NON_TER 1 1
SQ SEQUENCE 847 AA; 95944 MW; 15527E10825BB6CC CRC64;

Query Match 99.0%; Score 4053; DB 4; Length 847;
Best Local Similarity 99.9%; Pred. No. 2.8e-157;

Matches 815; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          9 EYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDY 68

Db          1 EYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDY 60

Qy         69 EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQL 128

Db         61 EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQL 120

Qy        129 RNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEEN 188

Db        121 RNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEEN 180

Qy        189 ISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLK 248

Db        181 ISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLK 240

Qy        249 TEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGL 308

Db        241 TEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGL 300

Qy        309 AKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQD 368

Db        301 AKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQD 360

Qy        369 TQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVD 428

Db        361 TQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVD 420

Qy        429 INGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLL 488

Db        421 INGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLL 480

Qy        489 EKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNET 548

Db        481 EKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNET 540

Qy        549 PNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPS 608

Db        541 PNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPS 600

Qy        609 PGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKD 668

Db        601 PGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKD 660

Qy        669 KEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMK 728

Db        661 KEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMK 720

Qy        729 LRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQ 788

Db        721 LRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQ 780

Qy        789 KLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824

Db        781 KLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSV 816
```

RESULT 3
Q921C5
ID Q921C5 PRELIMINARY; PRT; 820 AA.
AC Q921C5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bicaudal D protein.
GN BICD2 OR BICD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21376052; PubMed=11483508;
RA   Hoogenraad C.C., Akhmanova A., Howell S.A., Dortland B.R.,
RA   de Zeeuw C.I., Willemsen R., Visser P., Grosveld F., Galjart N.;
RT   "Mammalian Golgi-associated Bicaudal-D2 functions in the dynein-
RT   dynactin pathway by interacting with these complexes.";
RL   EMBO J. 20:4041-4054(2001).
DR   EMBL; AJ250106; CAC51393.1; -.
DR   MGD; MGI:1924145; Bicd2.
DR   InterPro; IPR003124; WH2.
DR   SMART; SM00246; WH2; 1.
SQ   SEQUENCE   820 AA;  93390 MW;  0C1A1754CD74DDE1 CRC64;

  Query Match           94.6%;  Score 3875;  DB 11;  Length 820;
  Best Local Similarity  94.8%;  Pred. No. 4.6e-150;
  Matches  783;  Conservative   13;  Mismatches   22;  Indels    8;  Gaps    2;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLTSVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDD--AAEPNNDAEA 298
              ||||||||||||||:|||||||||||||||||| |||||||||||  |||||||||||||
Db        241 EEALETLKTEREQKNNLRKELSHYMSINDSFYTSHLQVSLDGLKFSDDTVTAEPNNDAEA 300

Qy        299 LVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMERE 358
              ||||||| || |  |||||||:|:||||||||||||||||:|||||||||:|||||||||
Db        301 LVNGFEHSGLVKSSLDNKTSTPRKDGLAPPSPSLVSDLLSELHISEIQKLKQQLVQMERE 360

Qy        359 KAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDS 418
              | |||||||||||||  ||:|||| ||| |||||||||||||| ||||||:|||||||||
Db        361 KVGLLATLQDTQKQLEQARGTLSEQHEKVNRLTENLSALRRLQAGKERQTSLDNEKDRDS 420

Qy        419 HEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEG 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEG 480

Qy        479 QALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLY 538
              |||||||:||||||| ||||||| ||||||||||||||||||:|||||||||||||||||
Db        481 QALTEKISLLEKASHQDRELLAHLEKELKKVSDVAGETQGSLNVAQDELVTFSEELANLY 540

Qy        539 HHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRA 598
              |||||||||||||||||||| |      |||||| ||||||:||||||| | |||
Db        541 HHVCMCNNETPNRVMLDYYREGQGKA------GRTSPEGRGRRSPVLLPKGLLATEVGRA 594

Qy        599 DGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIAS 658
              ||||||:|||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 DGGTGDNSPSPSSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIAS 654

Qy        659 QELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENE 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 QELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENE 714

Qy        719 KAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTL 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 KAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTL 774

Qy        779 NSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||||||||||||||||||:||| | |||:|||
Db        775 NSLLRMAIQQKLALTQRLELLELDHEQTRRGRSKAASKAKPASPSL 820
```

PRT; 545 AA.
Created)
+ sequence update)
nnotation update)

```
RESULT 4
Q96LH2
ID   Q96LH2      PRELIMINARY;      PRT;   775 AA.      ata; Euteleostomi;
AC   Q96LH2;                                          e; Murinae; Mus.
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BA476B13.3 (KIAA0699 protein) (Fragment).
GN   BA476B13.3.
OS   Homo sapiens (Human).                                   H.,
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laird G.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137074; CAC88351.1; -.
DR   InterPro; IPR003124; WH2.
DR   SMART; SM00246; WH2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   775 AA;  87359 MW;  4345FC6B183718E2 CRC64;

  Query Match           90.3%;  Score 3696;  DB 4;  Length 775;
  Best Local Similarity  99.9%;  Pred. No. 8.1e-143;
  Matches  743;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         81 AFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENE 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENE 60

Qy        141 RLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQ 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQ 120

Qy        201 NQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKE 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKE 180

Qy        261 LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTP 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTP 240

Qy        321 KKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSL 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSL 300

Qy        381 SEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYH 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYH 360

Qy        441 VAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLA 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLA 420

Qy        501 RLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREG 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREG 480

Qy        561 QGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDP 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 QGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDP 540

Qy        621 RREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLK 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLK 600

Qy        681 SLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDA 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 SLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDA 660

Qy        741 ATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLE 800
```

```
Qy        617 ----------------LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
                              || |:||||||| |||||||||| |||| :|||| |::|
Db        599 TISPVITAPPSSPVLDTSDIRKEPMNIYNLNAIIRDQIKHLQKAVDRSLQLSRQRAAARE 658

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              | | :|||||||||||||||||||||||| ||| |||||||||||||||:|||||||
Db        659 LAPMIDKDKEALMEEILKLKSLLSTKREQIATLRAVLKANKQTAEVALANLKNKYENEKA 718

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              |||||| |||||||||||||||||||| |||||||||:|||||||||||||||||||:
Db        719 MVTETMTKLRNELKALKEDAATFSSLRTMFATRCDEYVTQLDEMQRQLAAAEDEKKTLNT 778

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||| || ||||:|| : |   |:|  :| :
Db        779 LLRMAIQQKLALTQRLEDLEFDHEQSRRSKGKLG-KSKIGSPKV 821


RESULT 7
Q91YP7
ID   Q91YP7      PRELIMINARY;      PRT;   835 AA.
AC   Q91YP7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to bicaudal D (Drosophila) homolog 1.
GN   BICD1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC016192; AAH16192.1; -.
DR   MGD; MGI:1101760; Bicd1.
DR   InterPro; IPR003124; WH2.
DR   SMART; SM00246; WH2; 1.
SQ   SEQUENCE   835 AA;  95896 MW;  33E8D83C56077C40 CRC64;

  Query Match           61.7%;  Score 2528;  DB 11;  Length 835;
  Best Local Similarity   64.0%;  Pred. No. 2.4e-95;
  Matches  527;  Conservative  110;  Mismatches  149;  Indels   38;  Gaps    8;

Qy         23 RAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEAF 82
              : |::||: || ||| |||||||||| |||||  || |::||| :|: :: |:|||:|||
Db         14 KTEIERLTKELTETTHEKIQAAEYGLVVLEEKLTLKQQYDELEAEYDGLKQELEQLREAF 73

Qy         83 GQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERL 142
              ||: : |:||| |||:|||:|:|||||| ||: |:||:| |||| | |:|| |:|||||
Db         74 GQSFSIHRKVAEDGETREETLLQESASKEAYYLNKILEMQNELKQSRAVVTNVQAENERL 133

Qy        143 ASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQ 202
              ::| ||||| |: ||:|| |::|:|:|||||||||||||:|||||||:||| || |:|||
Db        134 SAVVQELKENNEMVELQRIRMKDEIREYKFREARLLQDYTELEEENITLQKLVSTLKQNQ 193

Qy        203 VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELS 262
              ||:|||||||| |||| ||||||||||||:| |||||||||| ||||||:||||||
Db        194 VEYEGLKHEIKRFEEETVLLNSQLEDAIRLKEIAEHQLEEALETLKNEREQKNNLRKELS 253

Qy        263 HYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKK 322
              |::::|    ||: :|:|||||::| :||||| :  :||  || | ||  | :| | :|
Db        254 QYINLSD----SHISISVDGLKFAEDGSEPNNDDK--MNGHIHGPLGKLNGDYRTPTTRK 307

Qy        323 -EGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
               | | |     |||| |||||||||||||||:|:||||| ||| ||::| |||||:|:|:
Db        308 GESLHP-----VSDLFSELNISEIQKLKQQLIQVEREKAILLANLQESQTQLEHTKGALT 362

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
              || |:| ||||:::|:| || ||| :  || || |:| |:   ||||||| ||| ||| |
Db        363 EQHERVHRLTEHVNAMRGLQNSKEIKAELDCEKGRNSAEEAHDYEVDINGLEILECKYRV 422

Qy        442 AVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLAR 501
              || | :|:.::|||:  :      : ||| :||:: |   |:|: ||| ::  | :|
Db        423 AVTEVIDLKAEIKALKEKYNKSVENYTEEKTKYESKIQMYDEQVTNLEKTSKESGEKMAH 482

Qy        502 LEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQ 561
              :||||:|:: :| |   :|: |||||||||||| |||||:|||||||||||||: :
Db        483 MEKELQKMTGIANENHNTLNTAQDELVTFSEELAQLYHHVCLCNNETPNRVMLDYYRQSR 542

Qy        562 GGAGRTSPGGRTSP-EARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSP---- 616
                  | |   |: || || |:|: :|   |          :  |  |||
Db        543 ----VTRSGSLKGPDDPRGLLSPRLSRRGVSSPVESRTSSEPVSKENTETSKEPSPTKTP 598

Qy        617 ----------------LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
                              || |:||||||| |||||||||| |||| :|||| |::|
Db        599 TISPVITAPPSSPVLDTSDIRKEPMNIYNLNAIIRDQIKHLQKAVDRSLQLSRQRAAARE 658

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              | | :|||||||||||||||||||||||| ||| |||||||||||||||:|||||||
Db        659 LAPMIDKDKEALMEEILKLKSLLSTKREQIATLRAVLKANKQTAEVALANLKNKYENEKA 718

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              |||||| |||||||||||||||||||| |||||||||:|||||||||||||||||||:
Db        719 MVTETMTKLRNELKALKEDAATFSSLRAMFATRCDEYVTQLDEMQRQLAAAEDEKKTLNT 778

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||| || ||||:|| : |   |:|  :| :
Db        779 LLRMAIQQKLALTQRLEDLEFDHEQSRRSKGKLG-KSKIGSPKI 821


RESULT 8
O55206
ID   O55206      PRELIMINARY;      PRT;   545 AA.
AC   O55206;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Bicaudal-D (Fragment).
GN   BICD1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96079090; PubMed=8530100;
RA   Baens M., Aerssens J., Van Zand K., Cassiman J.J., Van Den Berghe H.,
RA   Marynen P.;
RT   "Isolation and regional assignment of chromosome 12p cDNAs.";
RL   Genomics 29:44-52(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98035884; PubMed=9367685;
RA   Baens M., Marynen P.;
RT   "A human homologue (BICD1) of the drosophila bicaudal-D gene.";
RL   Genomics 45:601-606(1997).
DR   EMBL; U90029; AAB94807.1; -.
DR   MGD; MGI:1101760; Bicd1.
FT   NON_TER       1      1
FT   NON_TER     545    545
SQ   SEQUENCE   545 AA;  61963 MW;  D4EB60DD6748C5F6 CRC64;

  Query Match           39.2%;  Score 1603.5;  DB 11;  Length 545;
  Best Local Similarity   61.2%;  Pred. No. 5.4e-58;
  Matches  343;  Conservative   69;  Mismatches  111;  Indels   37;  Gaps    7;

Qy        240 LEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEAL 299
              ||||||||| |||||:|||||| |::::|    ||: :|:|||||::| :||||| :
Db          1 LEEALETLKNEREQKNNLRKELSQYINLSD----SHISISVDGLKFAEDGSEPNNDDK-- 54

Qy        300 VNGFEHGGLAKLPLDNKTSTPKK-EGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMERE 358
              :|| || | ||  | :| | :| | | |     |||| |||||||||||||||:|:|||
Db         55 MNGHIHGPLGKLNGDYRTPTTRKGESLHP-----VSDLFSELNISEIQKLKQQLIQVERE 109
```

```
RESULT 11
Q94174
ID   Q94174      PRELIMINARY;      PRT;   869 AA.
AC   Q94174;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 99.7 kDa protein.
GN   C43G2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid C43G2.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U70848; AAB09108.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   869 AA;  99706 MW;  39A0A386BB21F093 CRC64;

  Query Match           21.0%;  Score 860;  DB 5;  Length 869;
  Best Local Similarity  30.3%;  Pred. No. 1.4e-27;
  Matches  255;  Conservative  144;  Mismatches  260;  Indels  182;  Gaps   20;

Qy        16 EAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEM 75
             |:: | || ::  |: :  :   :  :||  || :| :|  |: :  |:: : :  |:|:
Db         3 ESELEKLRQDIAILTEKYEQAKEDIHKAANAGLELLRQKEDLEKRLAEMQAELDLARTEI 62

Qy        76 EQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNT 135
             ::  :   :  : |::        ||||::||::||: |:::: :|: :||:    |
Db        63 DKTNQTLAEYRSQHQRSTRSELENEESLLEESSAKEEEYLQRIAKLEADLKKKEQELAEK 122

Qy       136 QSENERLASVAQELKEINQN--VEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQK 193
             : | | :   :  |||:   :| :| :|| ::|| | || ||: :|||||||||| |||
Db       123 KEELESIEK--KHSKEIDSGAALEDERRKLRAELKETKEREQRLISEYSELEEENIGLQK 180

Qy       194 QVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQ 253
             |: || :|||:| |: :  |||| | :    |:  |: |:::|||||| | : ||:|
Db       181 TVANLRGSQVEYESLRIDNNRLEETIEIMKMAAEEDEILRVIADKQLEEALLTAQQERDQ 240

Qy       254 KNSLRKEL------SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGG 307
             : ::::||       |  |:||  :                                |
Db       241 RLAMKRELEQTRNAEHISSLNDMLF--------------------------------G 266

Qy       308 LAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQ 367
             | :|  | :      | ||:|   ||| |||  |   |::    ::|  | ||   |:
Db       267 LERLGEDGE--------LPPPQPG-ASDLFSELQGSSDVKVR----ELEAAKEGLQEELK 313

Qy       368 DTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEV 427
               :|          | || | : |:  |             || |
Db       314 SREKIF----------IEFVTGLADTLNIHRP----------TNELD------------ 340

Qy       428 DINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSL 487
                 |  |  :   : |::: :          | :  |:: ||:   || |
Db       341 ----------YMHARQQKDVVLEKIQNI----------ARDTDRHDKEGE---EKRSG 375

Qy       488 LEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH------- 540
             : ||           :|: :  ||||    |: ||| ::  |::|   ||
Db       376 ILKA------------DLRTLVLVAGEKSAQLAAAQDAMIQVSDQLYQFYHQMTQNQGV 422

Qy       541 --------------VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPIL 585
                          :  | |   || |        |      |  :      |  |:
Db       423 QTEKSVQEIVKKLRLLARANAEDVPRVSL---------ADEGVESGTETDVNASRSIPLN 473

Qy       586 LPKGLLAP------EAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIK 639
               : ::||      |   |    ||            :| |:  :   | |:   : :|
Db       474 SDRLVIAPSFAKEIEKKLASVKIGDVLSE---------TDLRQRILTEGNAISETTESLK 524

Qy       640 HLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKA 699
               :   | ||:|      |  :   |   : |  |:  :||:||||||:||:||||||:
Db       525 KMIQVVKRTSEQ-----AFNQAVMASGAENEIEMQN-MKLRSLLSTKRDQISTLRTVLKS 578

Qy       700 NKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYIT 759
             || ||| || ::: |||:|| |: |   |:| ||| ||||||||:| ||||  | :|  :
Db       579 NKLTAESALTSMREKYESEKKMMMEINDKMRRELKQLKEDAATFASHRAMFTARGEELKS 638

Qy       760 QLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQT--RRGRAKAAPKT 817
             :::|:  :| | |:|||||| |||:|||||| |||||| :|:| ::   :|   :|. :
Db       639 KVEELSNELRANEEEKKTLNQLLRLAIQQKLTLTQRLEEVEVDRDRQVFKRSSTRAPTRE 698

Qy       818 K 818
             |
Db       699 K 699


RESULT 12
O43893
ID   O43893      PRELIMINARY;      PRT;   172 AA.
AC   O43893;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Bicaudal-D (Fragment).
GN   BICD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96079090; PubMed=8530100;
RA   Baens M., Aerssens J., Van Zand K., Cassiman J.J., Van Den Berghe H.,
RA   Marynen P.;
RT   "Isolation and regional assignment of chromosome 12p cDNAs.";
RL   Genomics 29:44-52(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98035884; PubMed=9367685;
RA   Baens M., Marynen P.;
RT   "A human homologue (BICD1) of the drosophila bicaudal-D gene.";
RL   Genomics 45:601-606(1997).
DR   EMBL; U90030; AAB94806.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   172 AA;  19336 MW;  13B71215ADA7A636 CRC64;

  Query Match           12.0%;  Score 491.5;  DB 4;  Length 172;
  Best Local Similarity  84.3%;  Pred. No. 2.2e-13;
  Matches  102;  Conservative    8;  Mismatches   10;  Indels    1;  Gaps    1;

Qy       704 AEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDE 763
             |||||||||:|||||||||||||| |||||||||||||||||||||||||||||:|||||
Db         1 AEVALANLKNKYENEKAMVTETMTKLRNELKALKEDAATFSSLRAMFATRCDEYVTQLDE 60

Qy       764 MQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPS 823
             |||||||||||||||:|||||||||||||||| || ||||:|| : |   |:|  :|
Db        61 MQRQLAAAEDEKKTLNTLLRMAIQQKLALTQRLEGLEFDHEQSRRSKGKLG-KSKIGSPK 119
```

```
Qy        824 L 824
              :
Db        120 V 120


RESULT 13
Q9H561
ID   Q9H561      PRELIMINARY;      PRT;    80 AA.
AC   Q9H561;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   BA526D8.1 (KIAA0699) (Fragment).
GN   BA526D8.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Babbage A.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL136981; CAC12725.1; -.
FT   NON_TER      80     80
SQ   SEQUENCE   80 AA;  9465 MW;  55CAB0B0C7BC4FEB CRC64;

  Query Match             9.7%;  Score 396;  DB 4;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-10;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKE 80
              ||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKE 80


RESULT 14
Q14789
ID   Q14789      PRELIMINARY;      PRT;  3259 AA.
AC   Q14789; Q14398;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GIANTIN (GCP372) (MACROGOLGIN) (Golgi autoantigen, golgin subfamily B,
DE   1).
GN   GOLGB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94187728; PubMed=7511208;
RA   Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA   Renz M.;
RT   "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT   protein (giantin) [retracted in Mol Cell Biol 1995 Jan;15(1):591].";
RL   Mol. Cell. Biol. 14:2564-2576(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94257116; PubMed=8198703;
RA   Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA   Renz M.;
RT   "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as
RT   target of antibodies in patients with rheumatic diseases and HIV
RT   infections.";
RL   J. Autoimmun. 7:67-91(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95100974; PubMed=7802676;
RA   Sohda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT   "Molecular cloning and sequence analysis of a human 372-kDA protein
RT   localized in the Golgi complex.";
RL   Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
CC   -!- FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES
CC       OF THE GOLGI COMPLEX.
CC   -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE
CC       SPLICING.
CC   -!- DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE
CC       AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.
DR   EMBL; X75304; CAA53052.1; -.
DR   EMBL; D25542; BAA05025.1; -.
DR   MIM; 602500; -.
KW   Golgi stack; Antigen; Coiled coil; Transmembrane;
KW   Alternative splicing.
FT   DOMAIN       1   3235       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  3236   3256       POTENTIAL.
FT   DOMAIN    3257   3259       LUMENAL (POTENTIAL).
FT   DOMAIN      48    110       COILED COIL (POTENTIAL).
FT   DOMAIN     127    223       COILED COIL (POTENTIAL).
FT   DOMAIN     238    448       COILED COIL (POTENTIAL).
FT   DOMAIN     460    526       COILED COIL (POTENTIAL).
FT   DOMAIN     545    593       COILED COIL (POTENTIAL).
FT   DOMAIN     677    956       COILED COIL (POTENTIAL).
FT   DOMAIN     969   1028       COILED COIL (POTENTIAL).
FT   DOMAIN    1062   1128       COILED COIL (POTENTIAL).
FT   DOMAIN    1154   1245       COILED COIL (POTENTIAL).
FT   DOMAIN    1301   1779       COILED COIL (POTENTIAL).
FT   DOMAIN    1828   2781       COILED COIL (POTENTIAL).
FT   DOMAIN    2797   2857       COILED COIL (POTENTIAL).
FT   DOMAIN    2872   2993       COILED COIL (POTENTIAL).
FT   DOMAIN    3026   3102       COILED COIL (POTENTIAL).
FT   DOMAIN    3133   3185       COILED COIL (POTENTIAL).
FT   DOMAIN    2420   2423       POLY-GLU.
FT   DOMAIN    2993   2996       POLY-SER.
FT   VARSPLIC     1     39       MISSING (IN REF. 3).
FT   VARSPLIC   215    215       A -> AQLSSM (IN REF. 3).
FT   CONFLICT  1765   1765       D -> G (IN REF. 3).
FT   CONFLICT  2950   2950       H -> D (IN REF. 3).
SQ   SEQUENCE  3259 AA;  376075 MW;  60376A20D8A178DD CRC64;

  Query Match             9.3%;  Score 382;  DB 4;  Length 3259;
  Best Local Similarity   21.8%;  Pred. No. 1.6e-07;
  Matches  205;  Conservative  179;  Mismatches  361;  Indels  196;  Gaps   34;

Qy          5 SEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLE------------ 52
              |  :|   |: | |      |||:|  :     :||| | |  :  ||
Db       2005 SHAKELQELLKEKQ-----QEVKQLQKDCIR-YQEKISALERTVKALEFVQTESQKDLEI 2058

Qy         53 --------EKHQLKLQFE-------------------------ELEVDYEAIRSEMEQ 77
                      :|: | | |                          ||: : |:::|:|:|
Db       2059 TKENLAQAVEHRKKAQAELASFKVLLDDTQSEAARVLADNLKLKKELQSNKESVKSQMKQ 2118

Qy         78 ----LKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLT 133
                  |:    ||   | |   : : : ::| :|    |:      : |  : |::: :
Db       2119 KDEDLERRLEQAEEKHLKEKKNMQEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQE 2178

Qy        134 NTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSE---LEEENIS 190
              |  |   :||:  : :  :  :     | |: |: |::: : :  :|  .|   |:|:| |
Db       2179 NLDSTVTQLAAFTKSMSSLQDD----RDRVIDEAKKWERKFSDAIQSKEEEIRLKEDNCS 2234

Qy        191 LQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKE--------ISERQLEE 242
              : |    |||  :  | ||  | ||| : :   |: :  ::|::          ::  | :
Db       2235 VLK--DQLRQMSIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLSQ 2292

Qy        243 ALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNG 302
                ||       :| | |  |   |: |    : |  ||:  |      |   : | ::
Db       2293 LEETRHLYHSSQNELAKLESELKSLKDQL--TDLSNSLEKCK------EQKGNLEGIIRQ 2344

Qy        303 FEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGL 362
              |    | :  ::| |  : |     |  | | |  |:|: | ||:   |   |
```

```
Db       2345 QE----ADIQ-NSKFSYEQLETDLQASRELTSRLHEEINMKE-QKIISLLSGKEEAIQVA 2398

Qy        363 LATL-QDTQKQLEHTRGSLSEQQE-------------KVTRLTENLSALRRLQASKERQ 407
              :| | |   |:::    ||:::|                |  :| | |  :::    :: |
Db       2399 IAELRQQHDKEIKELENLLSQEEEENIVLEEENKKAVDKTNQLMETLKTIKKENIQQKAQ 2458

Qy        408 --------TALDNEKDRDSHEDGDYYEVDINGPEILACKYHV---AVAEAGELREQLKAL 456
                     ::| |::||     ||| :::     |:  |  :   | ||  :|:|::: |
Db       2459 LDSFVKSMSSLQNDRDRIV---GDYQQLEERHLSIILEKDQLIQEAAAENNKLKEEIRGL 2515

Qy        457 RSTHEAREAQHAE---EKGRYEAE-GQALTEKVS----LLEKASRQDREL---LARLEKE 505
              ||  :   :::|:   |  :|  :  | :| | |    |||   :|::||    |:||::
Db       2516 RSHMDDLNSENAKLDAELIQYREDLNQVITIKDSQQKQLLEVQLQQNKELENKYAKLEEK 2575

Qy        506 LKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565
              ||:  :   : : |  :|    |:|: :|   :      :             : :|  |
Db       2576 LKESEEANEDLRRSFNALQEEKQDLSKEIESLKVSISQLTRQVT--------ALQEEGTLG 2628

Qy        566 RTSPGGRTSPEARGRRSPIL---------LPKGLLA--PEAGRADGGTGDSSPSPGSSLP 614
                   :   |   | | :          | : |:    || :  |   |
Db       2629 LYHAQLKVKEEEVHRLSALFSSSQKRIAELEEELVCVQKEAAKKVGEIEDKL-------- 2680

Qy        615 SPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALME 674
                  ::|  ::::   |:|::    : | :|  ||:|  :  ::     | |: : |
Db       2681 ------KKELKHLHHDAGIMRNE---TETAEERVAELARDLVEMEQKLLMVTKENKGLTA 2731

Qy        675 EILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYE---NEKAMVTETMMKLRN 731
              :|      :|:           |: ::  |   |  || ||:    | | : |  :  |
Db       2732 QIQSFGRSMSS-----------LQNSRDHANEELDELKRKYDASLKELAQLKEQGLLNRE 2780

Qy        732 ELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLA 791
                  | | | : :|         :  :: |::: :|| : :::   |:| |  :  |  :
Db       2781 RDALLSETAFSMNSTE-------ENSLSHLEKLNQQLLSKDEQLLHLSSQLEDSYNQVQS 2833

Qy        792 LTQRLELL--ELDH-----EQTRRG-----RAKAAPKTKPA 820
               :: :  |  | ||     |: |:      |: | | | ||
Db       2834 FSKAMASLQNERDHLWNELEKFRKSEEGKQRSAAQPSTSPA 2874


RESULT 15
Q9NJ22
ID   Q9NJ22      PRELIMINARY;      PRT;  1229 AA.
AC   Q9NJ22;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Myosin heavy chain catch (Smooth) muscle specific isoform
DE   (Fragment).
GN   MHC.
OS   Aequipecten irradians (Bay scallop).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pectinoida;
OC   Pectinoidea; Pectinidae; Argopecten.
OX   NCBI_TaxID=31199;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20090924; PubMed=10623517;
RA   Yamada A., Yoshio M., Oiwa K., Nyitray L.;
RT   "Catchin, a novel protein in molluscan catch muscles, is produced by
RT   alternative splicing from the myosin heavy chain gene.";
RL   J. Mol. Biol. 295:169-178(2000).
DR   EMBL; AF183909; AAF62392.1; -.
DR   HSSP; P24733; 1WDC.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR000533; Tropomyosin.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00194; TROPOMYOSIN.
DR   SMART; SM00015; IQ; 1.
DR   PROSITE; PS50096; IQ; 1.

FT   NON_TER       1      1
SQ   SEQUENCE   1229 AA;  142180 MW;  1764276CB904FDC9 CRC64;

  Query Match             8.9%;  Score 362.5;  DB 5;  Length 1229;
  Best Local Similarity   21.8%;  Pred. No. 3.4e-07;
  Matches  198;  Conservative  156;  Mismatches  350;  Indels  203;  Gaps   32;

Qy         13 LVMEAQPEWLRAEVK---RLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELE---- 65
              | :  | | :: ::|   ::  :||:| | | :  |  : :||:|: | || : ||
Db        118 LSIARQEEEMKEQLKQMDKMKEDLAKTERIKKELEEQNVTLLEQKNDLFLQLQTLEDSMG 177

Qy         66 ---------------------------------VDYEAIRSEME----QLKEAFGQAHT 87
                                                | | |: :||     ||:  |
Db        178 DQEERVEKLIMQKADFESQIKELEERLLDEEDAAADLEGIKKKMEADNANLKKDIGDLEN 237

Qy         88 NHKKVAADGESREE--SLIQESASKEQYYVRKVLELQTELKQLRNVLTNT----QSE--- 138
                :|   |   ::   | :|   |::  :: |   |  | | |      |    |:|
Db        238 TLQKAEQDKAHKDNQISTLQGEISQQDEHIGK---LNKEKKALEEANKKTSDSLQAEEDK 294

Qy        139 ----NERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQ 194
                  |:  | : | | |:  |:| :: ::| |::: |           ::|::  | |:
Db        295 CNHLNKLKAKLEQALDELEDNLEREK-KVRGDVEKAK----------RKVEQDLKSTQEN 343

Qy        195 VSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRL-----KEISERQ--LEEALETL 247
              |  | : : |   |:  ::| | |   |||:|||   |      ::| | |  :||  | |
Db        344 VEDLERVKRE---LEENVRRKEAEISSLNSKLEDEQNLVSQLQRKIKELQARIEELEEEL 400

Qy        248 KTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGG 307
              : ||  :  : |:            : |:  |: |     | |      |:  :|
Db        401 EAERNARAKVEKQ------------RAELNRELEELGERLDEAGGATSAQIELNKKREAE 448

Qy        308 LAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSEL--NISEIQKLKQQLMQMEREKAGLLAT 365
              | |:  | : :: : |           |  :|:   : ::||:|    : |:||  | :
Db        449 LLKIRRDLEEASLQHEAQISALRKKHQDAANEMADQVDQLQKVKS---KSEKEKQQLRSE 505

Qy        366 LQDTQKQLEHTRGSLSEQQEKVTRLTENLSALR-RLQASKERQTALDNEKDRDSHEDGDY 424
              ::| | |::|   :    :: : :    :| |  ||: |:      | ::| |   |: |
Db        506 VEDLQAQIQHISKNKGCSEKVMKQFESQMSDLNARLEDSQRSINELQSQKSRLQAENSDL 565

Qy        425 YEVDINGPEILACKYHVAV--AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALT 482
                     ::   :: |:|   |  :|  ||:      :|| :   |  : | : : :
Db        566 TR------QLEDAEHRVSVLSKEKSQLSSQLE------DARRSLEEETRARSKLQNEVRN 613

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDEL----VTFSEELANLY 538
                  :  | |:      :||:| :  |||    |  || | :|:      |  | ||
Db        614 MHADM--DAIRE------QLEEEQESKSDV----QRQLSKANNEIQQWRSKFESEGANRT 661

Qy        539 HHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPE---- 594
               :     |   | :|          |: |   :|:  |  : | :   |  |  |
Db        662 EEL-----EDQKRKLL----------GKLSEAEQTTEAANAKCSALEKAKSRLQQELEDM 706

Qy        595 ---------------AGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQI 638
                            ||   |      :  :|| | | : ::|       :  |:  |
Db        707 SIEVDRANASVNQMEKKQRAFDKTTAEWQAKVNSLQSELENSQKESRGYSAELYRIKASI 766

Qy        639 KHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLS---TKREQITTLRT 695
              :  | ::                 |: :: : | :||  |   ||       ::   |
Db        767 EEYQDSIG----------------ALRRENKNLADEIHDLTDQLSEGGRSTHELDKARR 809

Qy        696 VLKANKQTAEVALANLKSKYENEKAMVTETMMKL---RNEL-KALKEDAATFSSLRAMFA 751
               |:  |:  : ||   :   | |:| |    :::   |||: | ::|    | : |
Db        810 RLEMEKEELQAALEEAEGALEQEEAKVMRAQLEIATVRNEIDKRIQEKEEEFDNTRR--- 866

Qy        752 TRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRA 811
                    :  |: ||  | |   |          |:: |  | | :  ||:  : : ||:|
Db        867 ----NHQRALESMQASLEAEAKGKAD-------AMRIKKKLEQDINELEVALDASNRGKA 915

Qy        812 KAAPKTK 818
               :   |
Db        916 EMEKTVK 922


Search completed: July  1, 2003, 08:33:48
```

Job time : 41.5613 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 15.0815 Seconds
(without alignments)
2266.116 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1330	32.5	782	1	BICD_DROME	P16568 drosophila
2	358.5	8.8	1938	1	MYS_AEQIR	P24733 aequipecten
3	352.5	8.6	1790	1	USO1_YEAST	P25386 saccharomyc
4	352	8.6	1972	1	MYHB_HUMAN	P35749 homo sapien
5	346.5	8.5	2116	1	MYS2_DICDI	P08799 dictyosteli
6	346	8.4	1976	1	MYHA_HUMAN	P35580 homo sapien
7	345	8.4	1976	1	MYHA_RAT	Q9jlt0 rattus norv
8	344.5	8.4	1972	1	MYHB_RABIT	P35748 oryctolagus
9	341	8.3	1976	1	MYHA_BOVIN	Q27991 bos taurus
10	338.5	8.3	1978	1	MYHB_CHICK	P10587 gallus gall
11	337.5	8.2	1961	1	MYH9_RAT	Q62812 rattus norv
12	335.5	8.2	1972	1	MYHB_MOUSE	O08638 mus musculu
13	335	8.2	1960	1	MYH9_HUMAN	P35579 homo sapien
14	331	8.1	1962	1	MYSA_DROME	P05661 drosophila
15	331	8.1	2017	1	MYSN_DROME	Q99323 drosophila
16	328.5	8.0	1959	1	MYH9_CHICK	P14105 gallus gall
17	326.5	8.0	1130	1	YL17_CAEEL	Q11102 caenorhabdi
18	326.5	8.0	4473	1	PLE1_CRIGR	Q9ji55 cricetulus
19	325	7.9	2663	1	CENE_HUMAN	Q02224 homo sapien
20	323.5	7.9	1102	1	MYSC_CHICK	P29616 gallus gall
21	323.5	7.9	1957	1	YD86_SCHPO	Q10411 schizosacch
22	323	7.9	1433	1	REST_CHICK	O42184 gallus gall
23	321.5	7.9	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
24	321.5	7.9	1935	1	MYH7_HUMAN	P12883 homo sapien
25	320	7.8	1940	1	MYH3_CHICK	P02565 gallus gall
26	319.5	7.8	1940	1	MYH3_HUMAN	P11055 homo sapien
27	318.5	7.8	1935	1	MYH7_PIG	P79293 sus scrofa
28	317.5	7.8	1934	1	MYH7_MESAU	P13540 mesocricetu
29	317.5	7.8	1935	1	MYH7_RAT	P02564 rattus norv
30	317.5	7.8	1940	1	MYH3_RAT	P12847 rattus norv
31	317	7.7	1084	1	MYSS_RABIT	P02562 oryctolagus
32	315	7.7	1938	1	MYSS_CHICK	P13538 gallus gall
33	314.5	7.7	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
34	314.5	7.7	4684	1	PLE1_HUMAN	Q15149 homo sapien
35	314	7.7	4687	1	PLE1_RAT	P30427 rattus norv
36	310.5	7.6	1939	1	MYH1_HUMAN	P12882 homo sapien
37	310.5	7.6	1939	1	MYH4_HUMAN	Q9y623 homo sapien
38	308.5	7.5	1325	1	G160_MOUSE	P55937 mus musculu
39	308	7.5	879	1	MYSP_ONCVO	Q02171 onchocerca
40	308	7.5	1941	1	MYH2_HUMAN	Q9ukx2 homo sapien
41	305.5	7.5	1938	1	MYH4_RABIT	Q28641 oryctolagus
42	304.5	7.4	1939	1	MYH6_HUMAN	P13533 homo sapien
43	304	7.4	1875	1	MLP1_YEAST	Q02455 saccharomyc
44	303	7.4	3210	1	CENF_HUMAN	P49454 homo sapien
45	301	7.4	1938	1	MYH6_MOUSE	Q02566 mus musculu

ALIGNMENTS

RESULT 1
BICD_DROME
ID BICD_DROME STANDARD; PRT; 782 AA.
AC P16568; Q9VJD5;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytoskeleton-like bicaudal D protein.
GN BICD OR CG6605.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90075232; PubMed=2590944;
RA Wharton R.P., Struhl G.;
RT "Structure of the Drosophila BicaudalD protein and its role in
RT localizing the the posterior determinant nanos.";
RL Cell 59:881-892(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=90152340; PubMed=2576013;
RA Suter B., Romberg L.M., Steward R.;
RT "Bicaudal-D, a Drosophila gene involved in developmental asymmetry:
RT localized transcript accumulation in ovaries and sequence similarity
RT to myosin heavy chain tail domains.";
RL Genes Dev. 3:1957-1968(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR DIFFERENTIATION. IT
CC       MAY PLAY A ROLE IN LOCALIZING OF NANOS (A MATERNAL DETERMINANT)
CC       ACTIVITY IN OOCYTES. BICD MUTATIONS CAUSE NANOS MISLOCALIZATION
CC       AND THUS BICAUDAL DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: OOGENESIS.
CC   -!- SIMILARITY: OF C-TERMINAL HALF TO VARIOUS MYOSIN HEAVY CHAINS AND
CC       SOME INTERMEDIATE FILAMENT PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31684; AAA28393.1; -.
DR   EMBL; X51652; CAA35964.1; -.
DR   EMBL; AE003655; AAF53616.1; -.
DR   PIR; A33636; A33636.
DR   FlyBase; FBgn0000183; BicD.
KW   Coiled coil; Developmental protein.
FT   DOMAIN        3    263       COILED COIL (POTENTIAL).
FT   DOMAIN      319    477       COILED COIL (POTENTIAL).
FT   DOMAIN      601    746       COILED COIL (POTENTIAL).
FT   CONFLICT    296    296       A -> S (IN REF. 1).
FT   CONFLICT    318    318       L -> P (IN REF. 1).
FT   CONFLICT    477    477       H -> R (IN REF. 1).
SQ   SEQUENCE   782 AA;  88953 MW;  5A71776171DF58E6 CRC64;

  Query Match          32.5%;  Score 1330;  DB 1;  Length 782;
  Best Local Similarity  39.1%;  Pred. No. 2.2e-46;
  Matches  326;  Conservative  148;  Mismatches  247;  Indels  112;  Gaps   16;

Qy        22 LRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEA 81
             |: ||:||: || : :    |:|:|||::|||| |: : |||| |: | |::  :||
Db        18 LQMEVERLTRELDQVSSASAQSAQYGLSLLEEKSALQQKCEELETLYDNTRHELDITQEA 77

Qy        82 FGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENER 141
                : |: |    | :|::|: |||::|    :: :|: |||||: |   ::| :|
Db        78 LTKFQTSQKVTNKTGIEQEDALLNESAARETSLNLQIFDLENELKQLRHELERVRNERDR 137

Qy       142 LASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQN 201
             :    :      : |  | ||: ::|: |||| |:| :|||||||||||||||| || :
Db       138 MLQENSDFGRDKSDSEADRLRLKSELKDLKFRETRMLSEYSELEEENISLQKQVSSLRSS 197

                QVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKEL 261
                 |||| ||||:|| || | || |:::   ||:|:|:|:|||||||: ||| | :|:|||
                  CAKHEIRRLTEEVELLNQQVDELANLKKIAEKQMEEALETLQGEREAKYALKKEL 257

                  SHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPK 321
                      :|  |:| |      |:  | :|  |  |
                    RSNMEDNASNNSDGEE-----ENLALKRLEADLSTELKS 310

                  QKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
                     || || || | | |:: |  |: ::  |
                  ESMESEKTHLTANLREAQTSLDKSQNELQ 363

Qy       382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
             ::  |  :: || :|          :|  | | |
Db       364 NFMSRLALLAAHVDALVQL-----------KKQIDVKEQG------------------ 392

Qy       442 AVAEAG----ELREQLKALRS------THEAREAQHAEEKGRYEAEGQALTEKVSLL-EK 490
               | |    || :||:|| |      |  |:|    :       :|   |:  : | :
Db       393 --KEGGQKKDELEQQLRALISQYANWFTLSAKEIDGLKTDIAELQKGLNYTDATTTLRNE 450

Qy       491 ASRQDRELLARLEKELKKVSDV---------AGETQGSLSVAQDELVTFSEELANLYHHV 541
               :   :||| .:| |   |||          ||:: ||   |:  ||  |::|| ||| |
Db       451 VTNLKNKLLATEQKSLDLQSDVQTLTHISQNAGQSLGS---ARSTLVALSDDLAQLYHLV 507

Qy       542 CMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGG 601
             |  | ||| ||:||:  :                |: ::: :    :   ::    | ||
Db       508 CTVNGETPTRVLLDHKTDDMS----FENDSLTAIQSQFKSDVFIAKPQIVEDLQGLADS- 562

Qy       602 TGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQEL 661
                                    : |   :  : |||||:|: ||: | :::: :| | |
Db       563 ----------------------VEIKKYVDTVSDQIKYLKTAVEHTIDMNKHKIRS-EG 598

Qy       662 GPAVDK----DKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYEN 717
             | |::|    : | | |:|:|||||||| ||||| ||| |||:|||||||| ||||||||
Db       599 GDALEKVNTEEMEELQEQIVKLKSLLSVKREQIGTLRNVLKSNKQTAEVALTNLKSKYEN 658

Qy       718 EKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKT 777
             || :|::|| |||||: |||||||||||||||| ||:||:||:|:: ||| |||:||||
Db       659 EKIIVSDTMSKLRNELRLLKEDAATFSSLRAMFAARCEEYVTQVDDLNRQLEAAEEEKKT 718

Qy       778 LNSLLRMAIQQKLALTQRLELLELDHEQ--------TRRGRA-KAAPKTKPAT 821
             || |||:|:|||||||||| :|:| |          :|| : |::  |:|::
Db       719 LNQLLRLAVQQKLALTQRLEEMEMDREMRHVRRPMPAQRGTSGKSSFSTRPSS 771


RESULT 2
MYS_AEQIR
ID   MYS_AEQIR      STANDARD;      PRT;  1938 AA.
AC   P24733;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, striated muscle.
OS   Aequipecten irradians (Bay scallop).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pectinoida;
OC   Pectinoidea; Pectinidae; Argopecten.
OX   NCBI_TaxID=31199;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adductor muscle;
RX   MEDLINE=92011595; PubMed=1917970;
RA   Nyitray L., Goodwin E.B., Szent-Gyoergyi A.G.;
RT   "Complete primary structure of a scallop striated muscle myosin heavy
RT   chain. Sequence comparison with other heavy chains reveals regions
RT   that might be critical for regulation.";
RL   J. Biol. Chem. 266:18469-18476(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adductor muscle;
RX   MEDLINE=91088319; PubMed=2263488;
RA   Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT   "Nucleotide sequence of full length cDNA for a scallop striated
RT   muscle myosin heavy chain.";
RL   Nucleic Acids Res. 18:7158-7158(1990).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RX   MEDLINE=94173332; PubMed=8127365;
RA   Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
RA   Szent-Gyorgyi A.G., Cohen C.;
RT   "Structure of the regulatory domain of scallop myosin at 2.8-A
RT   resolution.";
RL   Nature 368:306-312(1994).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 17.8236 Seconds
(without alignments)
1360.244 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....:QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	329.5	8.0	2101	1	US-08-466-390-4	Sequence 4, Appli
2	329.5	8.0	2101	1	US-08-470-950-4	Sequence 4, Appli
3	329.5	8.0	2101	1	US-08-467-781-4	Sequence 4, Appli
4	329.5	8.0	2101	1	US-08-195-487-4	Sequence 4, Appli
5	329.5	8.0	2101	2	US-08-483-924-4	Sequence 4, Appli
6	329.5	8.0	2101	4	US-09-452-294-1	Sequence 1, Appli
7	329.5	8.0	2101	5	PCT-US93-06160-4	Sequence 4, Appli
8	310	7.6	885	2	US-08-533-306A-4	Sequence 4, Appli
9	310	7.6	885	2	US-08-742-923A-4	Sequence 4, Appli
10	305.5	7.5	816	2	US-08-533-306A-6	Sequence 6, Appli
11	305.5	7.5	816	2	US-08-742-923A-6	Sequence 6, Appli
12	304.5	7.4	1939	4	US-09-310-187A-1	Sequence 1, Appli
13	303	7.4	2482	1	US-08-328-254-6	Sequence 6, Appli
14	303	7.4	3248	1	US-08-353-700-1	Sequence 1, Appli
15	303	7.4	3248	5	PCT-US95-16216-1	Sequence 1, Appli
16	294.5	7.2	1886	4	US-08-938-105-3	Sequence 3, Appli
17	294	7.2	1388	4	US-09-572-191-2	Sequence 2, Appli
18	294	7.2	1388	4	US-09-723-262-2	Sequence 2, Appli
19	294	7.2	1388	4	US-09-723-219-2	Sequence 2, Appli
20	282.5	6.9	1312	2	US-08-592-126-148	Sequence 148, App
21	282.5	6.9	1312	2	US-08-687-080-51	Sequence 51, Appl
22	277	6.8	1211	4	US-09-134-001C-4820	Sequence 4820, Ap
23	264	6.4	1388	2	US-08-685-576-1	Sequence 1, Appli
24	261.5	6.4	1388	2	US-08-685-576-4	Sequence 4, Appli
25	253	6.2	1354	3	US-08-685-871-2	Sequence 2, Appli
26	249	6.1	976	4	US-09-104-324B-4	Sequence 4, Appli
27	242	5.9	1093	5	PCT-US93-03077-1	Sequence 1, Appli
28	241.5	5.9	1375	4	US-09-722-139-2	Sequence 2, Appli
29	241.5	5.9	1375	4	US-09-721-832-2	Sequence 2, Appli
30	241.5	5.9	1375	4	US-09-721-689-2	Sequence 2, Appli
31	235.5	5.8	683	6	5210183-3	Patent No. 5210183
32	234.5	5.7	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
33	234.5	5.7	1898	1	US-08-056-200-94	Sequence 94, Appl
34	234.5	5.7	1898	2	US-08-800-644-94	Sequence 94, Appl
35	226.5	5.5	955	2	US-08-428-414A-3	Sequence 3, Appli
36	226	5.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
37	223.5	5.5	1637	4	US-09-718-692-2	Sequence 2, Appli
38	223.5	5.5	1637	4	US-09-718-852-2	Sequence 2, Appli
39	223.5	5.5	1637	4	US-09-718-815-2	Sequence 2, Appli
40	222.5	5.4	1279	4	US-09-724-517-2	Sequence 2, Appli
41	222.5	5.4	1279	4	US-09-641-807A-2	Sequence 2, Appli
42	222.5	5.4	1279	4	US-09-723-096-2	Sequence 2, Appli
43	221.5	5.4	803	4	US-09-154-750A-85	Sequence 85, Appl
44	220.5	5.4	955	1	US-08-006-676B-1	Sequence 1, Appli
45	220.5	5.4	955	1	US-08-282-845-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,390
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-390-4

  Query Match             8.0%;  Score 329.5;  DB 1;  Length 2101;
  Best Local Similarity   19.7%;  Pred. No. 6.3e-13;
  Matches  195;  Conservative  190;  Mismatches  351;  Indels  255;  Gaps   37;

Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
              ||:::    |  |||  :  :|  ||:   ::|          ||||:: |: :  :||
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS--------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
```

```
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554
Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604
Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS-----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637
Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671
Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730
Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261


RESULT 2
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,950
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-470-950-4

  Query Match             8.0%;  Score 329.5;  DB 1;  Length 2101;
  Best Local Similarity   19.7%;  Pred. No. 6.3e-13;
  Matches  195;  Conservative  190;  Mismatches  351;  Indels  255;  Gaps   37;

Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS--------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
```

```
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554

Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604

Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS-----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637

Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671

Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730

Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787

Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818

Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261
```

RESULT 3
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-4

Query Match 8.0%; Score 329.5; DB 1; Length 2101;
Best Local Similarity 19.7%; Pred. No. 6.3e-13;
Matches 195; Conservative 190; Mismatches 351; Indels 255; Gaps 37;

```
Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66

Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS--------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112

Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168

Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227

Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267

Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315

Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375

Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421

Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469

Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509

Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554

Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604

Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS-----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637

Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671

Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730

Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787

Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818

Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261
```

```
RESULT 4
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/195,487
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/901,701
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-195-487-4

  Query Match             8.0%;  Score 329.5;  DB 1;  Length 2101;
  Best Local Similarity   19.7%;  Pred. No. 6.3e-13;
  Matches  195;  Conservative  190;  Mismatches  351;  Indels  255;  Gaps   37;

Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
              ||:::    |  |||  :  :|  ||:   ::|          ||||:: |: :  :||
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS--------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
                ::                ::| ||:       |: ::    ::| | |  |   :|
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
                : :    : | :|| :::|: ||    |:   :||:: :|     :  |   |  ::
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
                  |: : |:::    | |::  |:| :   :||:|:::|    :    ||:
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
                :   :| : :   : | |   |||           |  :| || :  : |
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
                ||  ||      :  : |          : | :  ::|:| |        |:|
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
                                    || :   |:   :  :: :||  |   ||  :: |:
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
              }:||| |::      |:  |    | |  |    : ::   : |::|   :
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
               |: :: :    |:|          | | :|  :| ::: | |  :|  : : |:
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
                         || | |   |   |||: :|     : |||     || | | : |::|
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554
               :    | :    ||  |:::   |:|:| |   |       || :| ::
Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604
               ::  | ||    ::    : |              |: ::  :|    | |: :
Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS-----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637
               :   | :                | : |:: | |  :   :|        :||
Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671
               :: |:||       : :| :  ::::| :|                  || :    ||
Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730
               |  |: ||:     ::||  :|   |:| :                |: | : |:  ||:  :|
Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
               : : |    :  :| :   | | |:  ::    || : |:|    | :  |||:  :::
Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
               :::::  | ::|| : |     |   |   |
Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261


RESULT 5
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,924
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
```

```
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-483-924-4

  Query Match             8.0%;  Score 329.5;  DB 2;  Length 2101;
  Best Local Similarity   19.7%;  Pred. No. 6.3e-13;
  Matches  195;  Conservative  190;  Mismatches  351;  Indels  255;  Gaps   37;

Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
              ||:::    |  |||  :  :|  ||:   ::|          ||||:: |: :  :||
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS-------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
                ::                ::| ||:       |: ::    ::| | |  |   :|
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
              : :    : | :|| :::|: ||    |:   :||:: :|     :  |   |  ::
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
                  |: : |:::    | |::  |:| :   :||:|:::::|    :    ||:
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
              :   :| : :   : | |   |||           |  :| || :  : |
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
                ||  ||      :  : |          : | :  ::|:| |        |:|
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
                                  || :   |:   :  :: :||  |   ||  :: |:
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
              |:||| |::       |:  |    | |  |    : ::   : |::|   :
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
               |: :: :    |:|           | | :|  :| ::: | |  :|  : : |:
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
                         || | |   |  |||: :|    : |||     || | | : |::|
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554
              :    | :    ||  |:::   |:|:| |   |        || :| ::
Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604
               ::  | ||    ::     : |             |: ::  :|    | |: :
Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS-----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637
              :   | :                 | : |:: | |    :   :|        :||
Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671
              :: |:||       : :| :  ::::| :|                    || :    ||
Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730
                  |  |: ||:     ::|| :|   |:| :           |: | : |:  ||:  :|
Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
                : : |    :  :| :   | | |:  ::    || : |:|    | :  |||:  :::
Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
              :::::  | ::|| : |      |   |   |
Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261


RESULT 6
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
;  APPLICANT:  Lelievre, Sophie
;  APPLICANT:  Bissell, Mina
;  TITLE OF INVENTION:  UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
;  TITLE OF INVENTION:  THERAPH AND DETECTION OF PROLIFERATIVE AND
;  TITLE OF INVENTION:  DIFFERENTIATION DISORDERS
;  FILE REFERENCE:  IB-1454- Sequence Submittal
; Patent No. 6287790
;  CURRENT APPLICATION NUMBER: US/09/452,294
;  CURRENT FILING DATE:  1999-11-30
;  PRIOR APPLICATION NUMBER:  60/110,420
;  PRIOR FILING DATE:  1998-11-30
;  NUMBER OF SEQ ID NOS:  1
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2101
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-452-294-1

  Query Match             8.0%;  Score 329.5;  DB 4;  Length 2101;
  Best Local Similarity   19.7%;  Pred. No. 6.3e-13;
  Matches  195;  Conservative  190;  Mismatches  351;  Indels  255;  Gaps   37;

Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
              ||:::    |  |||  :  :|  ||:   ::|          ||||:: |: :  :||
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS-------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
                ::                ::| ||:       |: ::    ::| | |  |   :|
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
              : :    : | :|| :::|: ||    |:   :||:: :|     :  |   |  ::
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
                  |: : |:::    | |::  |:| :   :||:|:::::|    :    ||:
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
              :   :| : :   : | |   |||           |  :| || :  : |
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
                ||  ||      :  : |          : | :  ::|:| |        |:|
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
                                  || :   |:   :  :: :||  |   ||  :: |:
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
              |:||| |::       |:  |    | |  |    : ::   : |::|   :
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
```

```
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
                          || | |    |   |||: :|     : |||     || | | : |::|
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554
               :    | :    ||  |:::   |:|:| |   |        || :| :: .
Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604
               ::  | ||    ::     : |             |: ::  :|    | |: :
Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637
               :   | :                | : |:: | |   :   :|        :||
Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671
               :: |:||       : :| :  ::::| :|                  || :    ||
Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730
               |  |: ||:     ::||  :|   |:| :           |: | : |:  ||:  :|
Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
               : : |    :  :| :   | | |:  ::    || : |:|    | :  |||:  :::
Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
               :::::  | ::|| : |     |   |   |
Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261
```

```
RESULT 7
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06160
;      FILING DATE:  19930621
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

PCT-US93-06160-4

Query Match 8.0%; Score 329.5; DB 5; Length 2101;
Best Local Similarity 19.7%; Pred. No. 6.3e-13;
Matches 195; Conservative 190; Mismatches 351; Indels 255; Gaps 37;

```
Qy          8 EEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQ-LKLQFEELEV 66
              ||:::    |  |||  :  :|  ||:   ::|           ||||:: |: :  :||
Db        346 EEHSK----ATQEWLEKQA-QLEKELSAALQDK--------KCLEEKNEILQGKLSQLEE 392

Qy         67 DYEAIRS--------------EMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQ 112
                ::                ::| ||:       |: ::    ::| | |  |   :|
Db        393 HLSQLQDNPPQEKGEVLGDVLQLETLKQEAATLAANNTQL----QARVEMLETERGQQEA 448

Qy        113 YYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ----NVEIQRGRLRDDIK 168
               : :    : | :|| :::|: ||    |:   :||:: :|     :  |   |  ::
Db        449 KLLAERGHFEEEKQQLSSLITDLQSSISNLSQAKEELEQASQAHGARLTAQVASLTSELT 508

Qy        169 EYKFREARLLQDYSELEEENISLQKQVS-VLRQNQVEFEGLKHEIKRLEEETEYLNSQLE 227
                    :  |: : |:::    | |::  |:| :   :||:|::::|    :    ||:
Db        509 TLNATIQQQDQELAGLKQQAKEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQQLK 568

Qy        228 DAIRLKEISERQLEEALETLKTERE----------QKNSLRKELSHYMSI--------- 267
               :   :| : :   : | |   |||          |  :| || :  : |
Db        569 EVAEKQEATRQDHAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVAN 628

Qy        268 --NDSFYTSHLHVSLDGLKFS----------DDAAEPNNDAEALVNGFEHGGLAKLPLDN 315
                 ||  ||      :  : |          : | :  ::|:| |        |:|
Db        629 EARDSAQTSVTQAQREKAELSRKVEELQACVETARQEQHEAQAQV--------AEL---- 676

Qy        316 KTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH 375
                                  || :   |:   :  :: :||  |   ||  :: |:
Db        677 ----------------------ELQLRSEQQKATEKERVAQEKDQLQEQLQALKESLKV 713

Qy        376 TRGSLSEQQ-------EKVTRLTENLSALRRLQASKERQTALDNEKDRDSHED------- 421
               |:||| |::       |:  |    | |  |    : ::   : |::|   :
Db        714 TKGSLEEEKRRAADALEEQQRCISELKAETRSLVEQHKRERKELEEERAGRKGLEARLLQ 773

Qy        422 -GDYYEVDINGPEIL-------ACKYHVAVAEAGELREQLKALRSTHEAREAQHAE---- 469
               |: :: :    |:|           | | :|  :| ::: | |  :|  : : |:
Db        774 LGEAHQAE---TEVLRRELAEAMAAQHTAESECEQLVKEVAAWRDGYEDSQQEEAQYGAM 830

Qy        470 ------------EKGRYEAEGQALTEKVSLLE-----KASRQDR---ELLARLEKELKKV 509
                         || | |    |   |||: :|     : |||     || | | : |::|
Db        831 FQEQLMTLKEECEKARQEL--QEAKEKVAGIESHSELQISRQQNKLAELHANLARALQQV 888

Qy        510 SDV---AGETQGSLSVAQDELVTFSEELANLYHHVCMC--NNETPNRVML---------- 554
               :    | :    ||  |:::   |:|:| |   |        || :| ::
Db        889 QEKEVRAQKLADDLSTLQEKMAATSKEVARLETLVRKAGEQQETASRELVKEPARAGDRQ 948

Qy        555 -DYYREGQGGAGRTSPGGRTSPEARGR---------RSPILLPKGLLAPEAGRADGGTGD 604
               ::  | ||    ::     : |             |: ::  :|    | |: :
Db        949 PEWLEEQQGRQFCSTQAALQAMEREAEQMGNELERLRAALMESQGQQQEERGQQEREVAR 1008

Qy        605 SSPSPGSS----------------LPSPLSDPRREPMNIYNLIAII------RDQ---- 637
               :   | :                | : |:: | |   :   :|        :||
Db       1009 LTQERGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHALTEKEGKDQELAK 1068

Qy        638 IKHLQAA-------VDRTTELSRQRIASQE------------------LGPAVDKDKEA 671
               :: |:|       : :| :  ::::| :|                  || :    ||
Db       1069 LRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL----EA 1124

Qy        672 LMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM-KLR 730
               |  |: ||:     ::||  :|   |:| :           |: | : |:  ||:  :|
Db       1125 LRAEVSKLEQQCQKQQEQADSLERSLEAER----------ASRAERDSAL--ETLQGQLE 1172

Qy        731 NELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
               : : |    :  :| :   | | |:  ::    || : |:|    | :  |||:  :::
Db       1173 EKAQELGHSQSALASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLI-SSLE 1231

Qy        788 QKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
               :::::  | ::|| : |     |   |   |
Db       1232 EEVSILNR-QVLEKEGESKELKRLVMAESEK 1261
```

```
RESULT 8
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-306A-4

  Query Match             7.6%;  Score 310;  DB 2;  Length 885;
  Best Local Similarity   22.1%;  Pred. No. 3.7e-12;
  Matches  198;  Conservative  121;  Mismatches  321;  Indels  254;  Gaps   33;

Qy          7 EEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELE- 65
              ||  |:  ::   : |  |   |: ||    : | |  |:      :| :|: | :||:
Db        161 EEMEAKANLDKNKQTLEKENADLAGELRVLGQAK-QEVEH------KKKKLEAQVQELQS 213

Qy         66 --VDYEAIRSEME--------------QLKEAFGQAHTNHKKVAADGESREESLIQESA 108
                | |  |:|:                 | || |:|    | ||:       | :|::
Db        214 KCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKLAKDVAS-----LSSQLQDTQ 268

Qy        109 SKEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIK 168
                |   |: | : |:|:||       : |   |      |  |  ||:|     |   :
Db        269 ELLQEETRQKLNVSTKLRQL-------EEERNSLQDQLDEEMEAKQNLERHISTLNIQLS 321

Qy        169 EYKFREARLLQDYSE----LEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNS 224
              : |    : |||::     |||     ||::  | |   ::|       :||:    |
Db        322 DSK----KKLQDFASTVEALEEGKKRFQKEIENLTQ---QYEEKAAAYDKLEKTKNRLQQ 374

Qy        225 QLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLK 284
              :|:| : :   ::|||   ||  : : :|  :  | :|                     |
Db        375 ELDDLV-VDLDNQRQLVSNLEKKQRKFDQLLAEEKNISS-------------------K 413

Qy        285 FSD-------DAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLL 337
              ::|       :| |    | :|    |    ||  |:      | |         : ||:
Db        414 YADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAE---------MEDLV 464

Qy        338 SE-----LNISEIQKLKQQL-MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLT 391
              |       |: |::| |: |  |||  |           |||     |   ::   ||
Db        465 SSKDDVGKNVHELEKSKRALETQMEEMKT----------QLEELEDELQASEDAKLRLE 513

Qy        392 ENLSAL-----RRLQASKERQTALDNEKDRDSHE----------------------DGDY 424
              |: ||     | |||  |:       :  | .||                      :||
Db        514 VNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNERALAAAAKKKLEGDL 573

Qy        425 YEVDINGPEILACKYHVAVAEAGELREQLKAL-RSTHEAR--------EAQHAEEKGR-Y 474
              ::::     :  :   |: :  :|: |:|   |   :||         |:  |:| :
Db        574 KDLELQADSAIKGR-EEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSL 632

Qy        475 EAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLS---VAQDELVTFS 531
              ||:   | | ::  |:| :|     | ||||     ::| |   |||     |||
Db        633 EADLMQLQEDLAAAERARKQ-----ADLEKE-----ELAEELASSLSGRNALQDEKRRLE 682

Qy        532 EELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLL 591
              :| |   :                | ||
Db        683 ARIAQLEEEL---------------EEEQGNM------------------------- 699

Qy        592 APEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTEL 651
                                      :||  |                |  | |   : ||
Db        700 ----------------------EAMSDRVR---------------KATQQAEQLSNEL 720

Qy        652 SRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANL 711
              : :|  :|       |:: |  :   : | | |   |    :::  |:     |  :| |
Db        721 ATERSTAQ-------KNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIAALEAKIAQL 773

Qy        712 KSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAM---FATRCDEYITQLDEMQRQL 768
              : : | |          |: : | |||        | |   :  : ::   :: :::|||
Db        774 EEQVEQEAREKQAATKSLKQKDKKLKEILLQVEDERKMAEQYKEQAEKGNARVKQLKRQL 833

Qy        769 AAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              ||:| : :|     | ::||   || || |        ||   | |:|   |
Db        834 EEAEEESQRIN-----ANRRKL---QR-ELDEATESNEAMGREVNALKSKLRGP 878


RESULT 9
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,923A
;      FILING DATE:  No. 5869611ember 1, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-742-923A-4

  Query Match             7.6%;  Score 310;  DB 2;  Length 885;
  Best Local Similarity   22.1%;  Pred. No. 3.7e-12;
  Matches  198;  Conservative  121;  Mismatches  321;  Indels  254;  Gaps   33;

Qy          7 EEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELE- 65
              ||  |:  ::   : |  |   |: ||    : | |  |:      :| :|: | :||:
Db        161 EEMEAKANLDKNKQTLEKENADLAGELRVLGQAK-QEVEH------KKKKLEAQVQELQS 213

Qy         66 --VDYEAIRSEME--------------QLKEAFGQAHTNHKKVAADGESREESLIQESA 108
                | |  |:|:                | || |:|    | ||:       | :|:::
Db        214 KCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKLAKDVAS-----LSSQLQDTQ 268

Qy        109 SKEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIK 168
              |  |: | : |:|:||       : |   |      |  |  ||:|     |   :
Db        269 ELLQEETRQKLNVSTKLRQL-------EEERNSLQDQLDEEMEAKQNLERHISTLNIQLS 321

Qy        169 EYKFREARLLQDYSE----LEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNS 224
              : |    : |||::     |||      ||::  | |   ::|       :||:    |
Db        322 DSK----KKLQDFASTVEALEEGKKRFQKEIENLTQ---QYEEKAAAYDKLEKTKNRLQQ 374

Qy        225 QLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLK 284
              :|:| : :   ::|||   ||  : : :|  :  | :|                     |
Db        375 ELDDLV-VDLDNQRQLVSNLEKKQRKFDQLLAEEKNISS--------------------K 413

Qy        285 FSD-------DAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLL 337
              ::|       :| |    | :|    |    ||  |:       | |          : ||:
Db        414 YADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAE---------MEDLV 464

Qy        338 SE-----LNISEIQKLKQQL-MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLT 391
              |       |: |::| |: |  |||  |            |||     |   ::  :||
Db        465 SSKDDVGKNVHELEKSKRALETQMEEMKT----------QLEELEDELQASEDAKLRLE 513

Qy        392 ENLSAL-----RRLQASKERQTALDNEKDRDSHE---------------------DGDY 424
               |: ||     | |||  |:       :  |  ||                     :||
Db        514 VNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNERALAAAAKKKLEGDL 573

Qy        425 YEVDINGPEILACKYHVAVAEAGELREQLKAL-RSTHEAR--------EAQHAEEKGR-Y 474
              :::::     :  :   |: :  :|: |:|   |   :||          |:  |:| :
Db        574 KDLELQADSAIKGR-EEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSL 632

Qy        475 EAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLS---VAQDELVTFS 531
              ||:   | | ::  |:| :|     | ||||     ::| |   |||     |||
Db        633 EADLMQLQEDLAAAERARKQ-----ADLEKE-----ELAEELASSLSGRNALQDEKRRLE 682

Qy        532 EELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLL 591
              :| |   :                | ||
Db        683 ARIAQLEEEL---------------EEEQGNM-------------------------- 699

Qy        592 APEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTEL 651
                                      :||  |                | | |   : ||
Db        700 ----------------------EAMSDRVR---------------KATQQAEQLSNEL 720

Qy        652 SRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANL 711
              : :|  :|       |:: |  :   : | | |   |    :::  |:     |  :| |
Db        721 ATERSTAQ-------KNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIAALEAKIAQL 773

Qy        712 KSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAM---FATRCDEYITQLDEMQRQL 768
              : : | |          |: : | |||        | |   :  : ::   :: :::|||
Db        774 EEQVEQEAREKQAATKSLKQKDKKLKEILLQVEDERKMAEQYKEQAEKGNARVKQLKRQL 833

Qy        769 AAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
                ||:| : :|     | ::||   || || |        ||   | |:|   |
Db        834 EEAEEESQRIN-----ANRRKL---QR-ELDEATESNEAMGREVNALKSKLRGP 878


RESULT 10
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  816 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-306A-6

  Query Match             7.5%;  Score 305.5;  DB 2;  Length 816;
  Best Local Similarity   22.1%;  Pred. No. 6.5e-12;
  Matches  197;  Conservative  122;  Mismatches  329;  Indels  245;  Gaps   34;

Qy          9 EYARLVMEAQPEWLRA---------------EVKRL-SHELAETTREKIQAAEYGLA--V 50
              ||  |  ||  .:|:|               :::||      |   |: |  |  ||
Db         83 EYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ-QEDALAQQA 141

Qy         51 LEEKHQLKLQFEELEVDY-EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESAS 109
               ||  :   :|||: :  : | : :|:| :     :|     |:| |  |   | :|::
Db        142 FEEARRRTREFEDRDRSHREEMENEVESVTGMLNEAEGKAIKLAKDVASL-SSQLQDTQE 200

Qy        110 KEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKE 169
                |   |: | : |:|:||       : |   |      |  |  ||:|     |   : :
Db        201 LLQEETRQKLNVSTKLRQL-------EEERNSLQDQLDEEMEAKQNLERHISTLNIQLSD 253

Qy        170 YKFREARLLQDYSE----LEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQ 225
              |    : |||::     |||      ||::  | |   ::|       :||:    | :
Db        254 SK----KKLQDFASTVEALEEGKKRFQKEIENLTQ---QYEEKAAAYDKLEKTKNRLQQE 306

Qy        226 LEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKF 285
              |:| : :   ::|||   ||  : : :|  :  | :|.                    |:
Db        307 LDDLV-VDLDNQRQLVSNLEKKQRKFDQLLAEEKNISS--------------------KY 345

Qy        286 SD-------DAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLS 338
              :|       :| |    | :|    |    ||  |:       | |          : ||:|
Db        346 ADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAE---------MEDLVS 396

Qy        339 E-----LNISEIQKLKQQL-MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTE 392
                      |: |::| |: |  |||  |            |||     |   ::  :||
Db        397 SKDDVGKNVHELEKSKRALETQMEEMKT----------QLEELEDELQASEDAKLRLEV 445
```

```
Qy        393 NLSAL-----RRLQASKERQTALDNEKDRDSHE----------------------DGDYY 425
              |: ||     | |||  |:      :  |  ||                       :||
Db        446 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNERALAAAAKKKLEGDLK 505

Qy        426 EVDINGPEILACKYHVAVAEAGELREQLKAL-RSTHEAR--------EAQHAEEKGR-YE 475
              ::::     :  :   |: :  :|: |:|   |   :||          |:  |:| :  |
Db        506 DLELQADSAIKGR-EEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSLE 564

Qy        476 AEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLS---VAQDELVTFSE 532
              |:   | | ::  |:| :|     | ||||     ::| |   |||     |||
Db        565 ADLMQLQEDLAAAERARKQ-----ADLEKE-----ELAEELASSLSGRNALQDEKRRLEA 614

Qy        533 ELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLA 592
              :| |   :             | ||
Db        615 RIAQLEEEL----------------EEEQGNM---------------------------- 630

Qy        593 PEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELS 652
                                    :||  |                 |  | |   : ||:
Db        631 ----------------------EAMSDRVR---------------KATQQAEQLSNELA 652

Qy        653 RQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLK 712
              :|  :|       |:: |  :   : | | |   |    ::: |:     |  :| |:
Db        653 TERSTAQ-------KNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIAALEAKIAQLE 705

Qy        713 SKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAM---FATRCDEYITQLDEMQRQLA 769
              : | |          |: : | |||  .      | |   :  : ::   :: :|||
Db        706 EQVEQEAREKQAATKSLKQKDKKLKEILLQVEDERKMAEQYKEQAEKGNARVKQLKRQLE 765

Qy        770 AAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              ||:| : :|    | ::||   || || |        ||   | |:|   |
Db        766 EAEEESQRIN-----ANRRKL---QR-ELDEATESNEAMGREVNALKSKLRGP 809
```

```
RESULT 11
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,923A
;      FILING DATE:  No. 5869611ember 1, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  816 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-742-923A-6

  Query Match             7.5%;  Score 305.5;  DB 2;  Length 816;
  Best Local Similarity   22.1%;  Pred. No. 6.5e-12;
  Matches  197;  Conservative  122;  Mismatches  329;  Indels  245;  Gaps   34;

Qy          9 EYARLVMEAQPEWLRA--------------EVKRL-SHELAETTREKIQAAEYGLA--V 50
              ||  |  ||   :|:|               :::||      |   |: |  |  ||
Db         83 EYVDLEREAGKVYLKAPMILNGVCVIWKGWIDLQRLDGMGCLEFDEERAQ-QEDALAQQA 141

Qy         51 LEEKHQLKLQFEELEVDY-EAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESAS 109
               ||  :  :||: :  : | : :|:| :     :|    |:| |  |   | :|::
Db        142 FEEARRRTREFEDRDRSHREEMENEVESVTGMLNEAEGKAIKLAKDVASL-SSQLQDTQE 200

Qy        110 KEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKE 169
                |   |: | : |:|:||       : |   |      |  |  ||:|     |  : :
Db        201 LLQEETRQKLNVSTKLRQL-------EEERNSLQDQLDEEMEAKQNLERHISTLNIQLSD 253

Qy        170 YKFREARLLQDYSE----LEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQ 225
              |    : |||::     |||     ||::  | |   ::|       :||:    | :
Db        254 SK----KKLQDFASTVEALEEGKKRFQKEIENLTQ---QYEEKAAAYDKLEKTKNRLQQE 306

Qy        226 LEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKF 285
              |:| : :   ::|||  ' ||  : : :|  :  | :|                    |:
Db        307 LDDLV-VDLDNQRQLVSNLEKKQRKFDQLLAEEKNISS-------------------KY 345

Qy        286 SD-------DAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLS 338
              :|       :| |    | :|    |    ||  |:      | |          : ||:|
Db        346 ADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAE---------MEDLVS 396

Qy        339 E-----LNISEIQKLKQQL-MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTE 392
                    |: |::| |: |  |||  |           |||     |   ::   ||
Db        397 SKDDVGKNVHELEKSKRALETQMEEMKT----------QLEELEDELQASEDAKLRLEV 445

Qy        393 NLSAL-----RRLQASKERQTALDNEKDRDSHE----------------------DGDYY 425
              |: ||     | |||  |:      :  |  ||                       :||
Db        446 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNERALAAAAKKKLEGDLK 505

Qy        426 EVDINGPEILACKYHVAVAEAGELREQLKAL-RSTHEAR--------EAQHAEEKGR-YE 475
              ::::     :  :   |: :  :|: |:|   |   :||          |:  |:| :  |
Db        506 DLELQADSAIKGR-EEAIKQLRKLQAQMKDFQRELEDARASRDEIFATAKENEKKAKSLE 564

Qy        476 AEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLS---VAQDELVTFSE 532
              |:   | | ::  |:| :|     | ||||     ::| |   |||     |||
Db        565 ADLMQLQEDLAAAERARKQ-----ADLEKE-----ELAEELASSLSGRNALQDEKRRLEA 614

Qy        533 ELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLA 592
              :| |   :             | ||
Db        615 RIAQLEEEL----------------EEEQGNM---------------------------- 630

Qy        593 PEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELS 652
                                    :||  |                 |  | |   : ||:
Db        631 ----------------------EAMSDRVR---------------KATQQAEQLSNELA 652

Qy        653 RQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLK 712
              :|  :|       |:: |  :   : | | |   |    ::: |:     |  :| |:
Db        653 TERSTAQ-------KNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIAALEAKIAQLE 705

Qy        713 SKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAM---FATRCDEYITQLDEMQRQLA 769
              : | |          |: : | |||         | |   :  : ::   :: :|||
Db        706 EQVEQEAREKQAATKSLKQKDKKLKEILLQVEDERKMAEQYKEQAEKGNARVKQLKRQLE 765

Qy        770 AAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              ||:| : :|     | ::||   || || |        ||   | |:|   |
Db        766 EAEEESQRIN-----ANRRKL---QR-ELDEATESNEAMGREVNALKSKLRGP 809
```

```
RESULT 12
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
```

```
; GENERAL INFORMATION:
;  APPLICANT:  Benichou, Gilles
;  APPLICANT:  Fedoseyeva, Eugenia
;  TITLE OF INVENTION:  Involvement of Autoantigens in Cardiac
;  TITLE OF INVENTION:  Graft Rejection
;  FILE REFERENCE:  UCSF-090
;  CURRENT APPLICATION NUMBER: US/09/310,187A
;  CURRENT FILING DATE:  1999-05-12
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-310-187A-1

  Query Match             7.4%;  Score 304.5;  DB 4;  Length 1939;
  Best Local Similarity   20.6%;  Pred. No. 2.3e-11;
  Matches  194;  Conservative  158;  Mismatches  363;  Indels  227;  Gaps   33;

Qy          7 EEEYARLVMEAQPEWLRAEVKRLSHELAETTREK------IQAAEYGLAVLEEK-HQL-- 57
                :||: |:  :   |    |  | |  ::    :||     :|| :  |    ||: ||
Db        855 KEEFGRI--KETLEKSEARRKELEEKMVSLLQEKNDLQLQVQAEQDNLNDAEERCDQLIK 912

Qy         58 -KLQFE------------------ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAAD--- 95
               |:| |                  ||       :  |  :||:          || :
Db        913 NKIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHA 972

Qy         96 GESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQN 155
               |:: ::| :| |  ::    :     | :    :   :  | : | | ::: |:::   :: |
Db        973 TENKVKNLTEEMAGLDEIIAKLTKEKKALQEAHQQALDDLQVEEDKVNSLSKSKVKLEQQ 1032

Qy        156 VEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVEFEG----LKHE 211
               |:   | |     :| | |          :||     |:  : : ::: :: |      |: :
Db       1033 VDDLEGSLE---QEKKVR--------MDLERAKRKLEGDLKLTQESIMDLENDKLQLEEK 1081

Qy        212 IKRLEEETEYLNSQLED----AIRL-KEISERQ-----LEEALETLKTEREQKNSLRKEL 261
               :|: | :    ||::||     |::| |:: | |     ||| ||   :| |  :    || :|
Db       1082 LKKKEFDINQQNSKIEDEQALALQLQKKLKENQARIEELEEELEAERTARAKVEKLRSDL 1141

Qy        262 SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPK 321
               |                      |: :     :|          : :|         |:  | : :| :
Db       1142 SR---------------ELEEISERLEEAGGATSVQIEMNKKREAEFQKMRRDLEEATLQ 1186

Qy        322 KEGLAPPSPSLVSDLLSEL--NISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEH---T 376
               |  |       :| ::||   |  :|::||:|   |:||:       | |       :|
Db       1187 HEATAAALRKKHADSVAELGEQIDNLQRVKQKL---EKEKSEFKLELDDVTSNMEQIIKA 1243

Qy        377 RGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILA 436
               : :|     |||:|  |: :    |::  : :::   |       |       :
Db       1244 KANL----EKVSRTLEDQANEYRVKLEEAQRSLNDFTTQRAKLQ--------------- 1283

Qy        437 CKYHVAVAEAGELREQLKALRSTHEAREA--------------QHAEEKGRYEAEGQALT 482
                        | |||  ||           | :||                |   :  |  :  |  ||:|
Db       1284 -------TENGELARQL-------EEKEALISQLTRGKLSYTQQMEDLKRQLEEEGKAKN 1329

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTF------------ 530
               |: |  |  |  :|   | ::  ::  ::     |  |   ||| |    |:   :
Db       1330 ALAHALQSA-RHDCDL---LREQYEEETEAKAELQRVLSKANSEVAQWRTKYETDAIQRT 1385

Qy        531 ----------------SEELANLYHHVCMCNNETPNR-------VMLDYYREGQGGAGRT 567
                                 :||       :       :| :|         :|:|    |        |
Db       1386 EELEEAKKKLAQRLQDAEEAVEAVNAKCSSLEKTKHRLQNEIEDLMVDVERSNAAAA--- 1442

Qy        568 SPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNI 627
                  : : : |      :| :        |   :                 | | |    :  |        :
Db       1443 ------ALDKKQRNFDKILAEWKQKYEESQ-------------SELESSQKEARSLSTEL 1483

Qy        628 YNLIAIIRDQIKHLQAAVDRTTELSRQRIA--SQELGPA------VDKDKEALMEEILKL 679
               : |         : ::||:      |   :  :: |:   :::||         ::| :: |   | |:|
Db       1484 FKLKNAYEESLEHLE-TFKRENKNLQEEISDLTEQLGEGGKNVHELEKVRKQLEVEKLEL 1542

Qy        680 KSLL-----STKREQITTLRTVLKANKQTAEV--ALANLKSKYENEKAMVTETMMKLRNE 732
               :| |      | : |:   ||  |: |:   ||    |:   : |      :  |  |         :  |:
Db       1543 QSALEEAEASLEHEEGKILRAQLEFNQIKAEIERKLAEKDEEMEQAKRNHQRVVDSLQTS 1602

Qy        733 LKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMA------- 785
               | |              ::           :|    ||       |     |  | |:  :  |:| |||:
Db       1603 LDAETRSRNEVLRVKKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLQSLLKDTQIQLDDA 1662

Qy        786 ------IQQKLALTQR--------LELLELDHEQTRRGRAKA 813
                     ::: :|: :|          || |     ||| |  |   |
Db       1663 VRANDDLKENIAIVERRNNLLQAELEELRAVVEQTERSRKLA 1704

RESULT 13
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
;  GENERAL INFORMATION:
;    APPLICANT:  Zhu, Xueliang
;    APPLICANT:  Lee, Wen-Hwa
;    TITLE OF INVENTION:  A No. 5710022el Nuclear Mitotic Phosphoprotein
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,254
;      FILING DATE:  24-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/141,239
;      FILING DATE:  22-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-CJ 1191
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2482 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-328-254-6

  Query Match             7.4%;  Score 303;  DB 1;  Length 2482;
  Best Local Similarity   20.6%;  Pred. No. 4e-11;
  Matches  195;  Conservative  166;  Mismatches  316;  Indels  270;  Gaps   38;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSH---ELAETTRE------KIQAAEYGLAVL 51
                :|: ||    |  :|    : |: |:  |   :| |  ||          |:: |  | :
Db       1475 LSSTQEEVHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQMS 1534

Qy         52 EEKHQ-----------------LKLQFEEL-------EVDYEAIRSEME----QLKEAFGQA 85
                ||  :                || | ||:              |:|    :|||  |     |::| ||
Db       1535 EENQELVILDAENSKAEVETLKTQIEEMARSLKVFELDLVTLRSEKENLTKQIQEKQGQL 1594

Qy         86 HTNHKKVAADGESREESLIQESASKEQYYVR-------KVLELQTELKQLRNVLTNTQSE 138
                  |  :       |   :||::|    |||   ::             |  || :||:|   :       :
Db       1595 SELDKLL-----SSFKSLLEE---KEQAEIQIKEESKTAVEMLQNQLKELNEAVAALCGD 1646

Qy        139 NERLASVAQEL-------KEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEE----- 186
                 | : :  | |               ::  ::|  | || | |        ::  :||   | |
```

```
Db       1647 QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK---KQLCVLQQLKESEHHADLL 1702

Qy        187 ----ENISLQKQVSVLRQNQ----VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISER 238
                  ||  |:::: : | ||    :| |  | |:: |: : | :   |           |
Db       1703 KGRVEN--LERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSL----------R 1749

Qy        239 QLEEALETLKTEREQ-KNSLRKE---LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN 294
              ||  : |:::|:|   | |:|| :|    || ||
Db       1750 GLELDVVTIRSEKENLTNELQKEQERISELEIINSSF---------------------- 1786

Qy        295 DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELN------------ 341
                    |  :     |: :  |:||         :  ::   | |||
Db       1787 -----ENILQEKEQEKVQMKEKSST---------AMEMLQTQLKELNERVAALHNDQEAC 1832

Qy        342 ISEIQKLKQQLMQMEREKAGLLATLQDTQ-------------KQLEHTRGSLSEQQEKV 387
               :: | |  |:  :| ||| ||  | : :              :::|  :  | :: |::
Db       1833 KAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVNGLIQEVEDGKQKLEKKDEEI 1892

Qy        388 TRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAG 447
              :||   :    :|  ||  |   :::  :: : :     |:      |  |  |  ::
Db       1893 SRLKNQIQDQEQL-VSKLSQVEGEHQLWKEQNLELRNLTVE------LEQKIQVLQSKNA 1945

Qy        448 ELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS-LLEKASRQDRELLARLEKEL 506
               |:: |: |:|:::  |     |    : : :  |||: :  | :   ||:     :|
Db       1946 SLQDTLEVLQSSYKNLE----NELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTA 2001

Qy        507 KKVSDVAGETQ---GSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGG 563
              :   :::||     | | :  :|: :  ::|  |     : |:|   :  ||  : |
Db       2002 ELQEELSGEKNRLAGELQLLLEEIKSSKDQLKEL----TLENSEL--KKSLDCMHKDQ-- 2053

Qy        564 AGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE 623
                        |  |:     :    |   || :
Db       2054 -----------VEKEGKVREEIAEYQLRLHEAEKKHQ---------------------- 2079

Qy        624 PMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL 683
                  |:: |  |  :  :    :  |::: |:|         | |  : |::  |
Db       2080 --------ALLLDTNKQYEVEI----QTYREKLTSKE---------ECLSSQKLEIDLLK 2118

Qy        684 STKREQITTLRTVLKANKQTAEVALANLK----SKYENEKA-----MVTETMMKLRNELK 734
              |:| |   :|:   :  ::  :  :|||     | |||:|     :: ::  :|  | :
Db       2119 SSKEELNNSLKATTQILEELKKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLEEEKE 2178

Qy        735 ALKEDAATFSSLR------AMFATRCDEYITQLDEMQRQLAAAEDEKKTLN--------- 779
               |::: :   : :       :  |: ||  |:: |::  |     |:||
Db       2179 ILQKELSQLQAAQEKQKTGTVMDTKVDELTTEIKELKETL-----EEKTKEADEYLDKYC 2233

Qy        780 SLL----RMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              |||    ::   :::  ||   |     :|  ||    :|   | |
Db       2234 SLLISHEKLEKAKEMLETQVAHLCSQQSKQDSRG----SPLLGPVVP 2276


RESULT 14
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
;  GENERAL INFORMATION:
;    APPLICANT:  YEN, TIMOTHY J.
;    APPLICANT:  RATTNER, JEROME B.
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING A
;    TITLE OF INVENTION:  TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DANN, DORFMAN, HERRELL AND SKILLMAN
;      STREET:  1601 MARKET STREET, SUITE 720
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,700
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REED, JANET E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3248 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
US-08-353-700-1

  Query Match             7.4%;  Score 303;  DB 1;  Length 3248;
  Best Local Similarity   20.6%;  Pred. No. 5.7e-11;
  Matches  195;  Conservative  166;  Mismatches  316;  Indels  270;  Gaps   38;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSH---ELAETTRE------KIQAAEYGLAVL 51
              :|:  ||     |  :|     : |: |:  |   :| |  ||      |::  |  | :
Db       2203 LSSTQEEVHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQMS 2262

Qy         52 EEKHQ---------------LKLQFEEL-------EVDYEAIRSEME----QLKEAFGQA 85
              ||  :                || | ||:       |:|   :||| |    |::|  ||
Db       2263 EENQELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQL 2322

Qy         86 HTNHKKVAADGESREESLIQESASKEQYYVR-------KVLELQTELKQLRNVLTNTQSE 138
                 | :     |  :||::|   |||  ::         |  || :||:|   :     :
Db       2323 SELDKLL-----SSFKSLLEE---KEQAEIQIKEESKTAVEMLQNQLKELNEAVAALCGD 2374

Qy        139 NERLASVAQEL-------KEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEE----- 186
              | : :  | |         ::  ::|  | ||  | |    ::  :||   | |
Db       2375 QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK----KQLCVLQQLKESEHHADLL 2430

Qy        187 ----ENISLQKQVSVLRQNQ----VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISER 238
                  ||  |:::: : | ||    :| |  | |:: |: : | :   |           |
Db       2431 KGRVEN--LERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSL----------R 2477

Qy        239 QLEEALETLKTEREQ-KNSLRKE---LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN 294
              ||  : |:::|:|   | |:|| :|    || ||
Db       2478 GLELDVVTIRSEKENLTNELQKEQERISELEIINSSF---------------------- 2514

Qy        295 DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELN------------ 341
                    |  :     |: :  |:||         :  ::   | |||
Db       2515 -----ENILQEKEQEKVQMKEKSST---------AMEMLQTQLKELNERVAALHNDQEAC 2560

Qy        342 ISEIQKLKQQLMQMEREKAGLLATLQDTQ-------------KQLEHTRGSLSEQQEKV 387
               :: | |  |:  :| ||| ||  | : :              :::|  :  | :: |::
Db       2561 KAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEKKDEEI 2620

Qy        388 TRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAG 447
              :||   :    :|  ||  |   :::  :: : :     |:      |  |  |  ::
Db       2621 SRLKNQIQDQEQL-VSKLSQVEGEHQLWKEQNLELRNLTVE------LEQKIQVLQSKNA 2673

Qy        448 ELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS-LLEKASRQDRELLARLEKEL 506
               |:: |: |:|:::  |     |    : : :  |||: :  | :   ||:     :|
Db       2674 SLQDTLEVLQSSYKNLE----NELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTA 2729

Qy        507 KKVSDVAGETQ---GSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGG 563
              :   :::||     | | :  :|: :  ::|  |     : |:|   :  ||  : |
Db       2730 ELQEELSGEKNRLAGELQLLLEEIKSSKDQLKEL----TLENSEL--KKSLDCMHKDQ-- 2781

Qy        564 AGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE 623
```

```
                     |  |:    :    |   || :
Db       2782 -----------VEKEGKVREEIAEYQLRLHEAEKKHQ---------------------- 2807

Qy        624 PMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL 683
                  |::  |  |  :  :    :  |::: |:|         | |  : |::  |
Db       2808 --------ALLLDTNKQYEVEI----QTYREKLTSKE---------ECLSSQKLEIDLLK 2846

Qy        684 STKREQITTLRTVLKANKQTAEVALANLK----SKYENEKA-----MVTETMMKLRNELK 734
              |:| |   :|:   :  ::  :  : |||      | |||:|     :: ::  :|  | :
Db       2847 SSKEELNNSLKATTQILEELKKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLEEEKE 2906

Qy        735 ALKEDAATFSSLR------AMFATRCDEYITQLDEMQRQLAAAEDEKKTLN--------- 779
                |::: :   : :        :  |: ||  |:: |::  |     |:||
Db       2907 ILQKELSQLQAAQEKQKTGTVMDTKVDELTTEIKELKETL-----EEKTKEADEYLDKYC 2961

Qy        780 SLL----RMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              |||    ::   :::  ||   |      :|  ||    :|   |  |
Db       2962 SLLISHEKLEKAKEMLETQVAHLCSQQSKQDSRG----SPLLGPVVP 3004


RESULT 15
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
;  GENERAL INFORMATION:
;    APPLICANT:  Yen, Timothy J.
;    APPLICANT:  Rattner, Jerome B.
;    TITLE OF INVENTION:  Nucleic Acid Encoding a Transiently
;    TITLE OF INVENTION:  Expressed Kinetochore Protein, and Methods of Use
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street   Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16216
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/353,700
;      FILING DATE:  09-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3248 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-16216-1

  Query Match             7.4%;  Score 303;  DB 5;  Length 3248;
  Best Local Similarity   20.6%;  Pred. No. 5.7e-11;
  Matches  195;  Conservative  166;  Mismatches  316;  Indels  270;  Gaps   38;

Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSH---ELAETTRE------KIQAAEYGLAVL 51
              :|:  ||    |  :|    : |: |:  |   :| |  ||      |::  |  | :
Db       2203 LSSTQEEVHQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQMS 2262

Qy         52 EEKHQ---------------LKLQFEEL-------EVDYEAIRSEME----QLKEAFGQA 85
              ||  :                || | ||:      |:|   :||| |    |::|  ||
Db       2263 EENQELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQL 2322

Qy         86 HTNHKKVAADGESREESLIQESASKEQYYVR-------KVLELQTELKQLRNVLTNTQSE 138
                 | :     |  :||::|   |||  ::         |  || :||:|   :     :
Db       2323 SELDKLL-----SSFKSLLEE---KEQAEIQIKEESKTAVEMLQNQLKELNEAVAALCGD 2374

Qy        139 NERLASVAQEL-------KEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEE----- 186
               | : :  | |         ::  ::|  | ||  | |    ::  :||   | |
Db       2375 QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK----KQLCVLQQLKESEHHADLL 2430

Qy        187 ----ENISLQKQVSVLRQNQ----VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISER 238
                  ||  |:::: : | ||     :| |  | |:: |: : | :   |           |
Db       2431 KGRVEN--LERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSL-----------R 2477

Qy        239 QLEEEALETLKTEREQ-KNSLRKE---LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN 294
               ||  : |:::|:|   | |:||   :|    || ||
Db       2478 GLELDVVTIRSEKENLTNELQKEQERISELEIINSSF---------------------- 2514

Qy        295 DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELN------------ 341
                  |  :     |: :  |:||         :  ::   | |||
Db       2515 -----ENILQEKEQEKVQMKEKSST---------AMEMLQTQLKELNERVAALHNDQEAC 2560

Qy        342 ISEIQKLKQQLMQMEREKAGLLATLQDTQ-------------KQLEHTRGSLSEQQEKV 387
               :: | |  |:  :| ||| ||  | : :              :::|  :  | :: |::
Db       2561 KAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEKKDEEI 2620

Qy        388 TRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAG 447
               :||  :   :|  ||  |   :::  ::: : :     |:      |  |  |  ::
Db       2621 SRLKNQIQDQEQL-VSKLSQVEGEHQLWKEQNLELRNLTVE------LEQKIQVLQSKNA 2673

Qy        448 ELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVS-LLEKASRQDRELLARLEKEL 506
               |:: |: |:|:::  |     |    : :  :  |||: :  | :   ||:    :|
Db       2674 SLQDTLEVLQSSYKNLE----NELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTA 2729

Qy        507 KKVSDVAGETQ---GSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGG 563
               :  :::||     | | :  :|: :  ::|  |     : |:|   :  ||   : |
Db       2730 ELQEELSGEKNRLAGELQLLLEEIKSSKDQLKEL----TLENSEL--KKSLDCMHKDQ-- 2781

Qy        564 AGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE 623
                    |  |:    :    |   || :
Db       2782 -----------VEKEGKVREEIAEYQLRLHEAEKKHQ---------------------- 2807

Qy        624 PMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL 683
                  |::  |  |  :  :    :  |::: |:|         | |  : |::  |
Db       2808 --------ALLLDTNKQYEVEI----QTYREKLTSKE---------ECLSSQKLEIDLLK 2846

Qy        684 STKREQITTLRTVLKANKQTAEVALANLK----SKYENEKA-----MVTETMMKLRNELK 734
              |:| |   :|:   :  ::  :  : |||      | |||:|     :: ::  :|  | :
Db       2847 SSKEELNNSLKATTQILEELKKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLEEEKE 2906

Qy        735 ALKEDAATFSSLR------AMFATRCDEYITQLDEMQRQLAAAEDEKKTLN--------- 779
                |::: :   : :        :  |: ||  |:: |::  |     |:||
Db       2907 ILQKELSQLQAAQEKQKTGTVMDTKVDELTTEIKELKETL-----EEKTKEADEYLDKYC 2961

Qy        780 SLL----RMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATP 822
              |||    ::   :::  ||   |      :|  ||    :|   |  |
Db       2962 SLLISHEKLEKAKEMLETQVAHLCSQQSKQDSRG----SPLLGPVVP 3004


Search completed: July  1, 2003, 08:35:58
Job time : 25.8236 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:33:58 ; Search time 126.593 Seconds
(without alignments)
713.694 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters: 424699

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4095	100.0	824	9	US-09-884-001-2	Sequence 2, Appli
2	2536	61.9	821	9	US-09-884-001-18	Sequence 18, Appl
3	2536	61.9	975	9	US-10-097-340-31	Sequence 31, Appl
4	420	10.3	170	9	US-09-746-783-16	Sequence 16, Appl
5	352	8.6	1938	9	US-10-171-311-164	Sequence 164, App
6	352	8.6	1945	9	US-09-927-597-2	Sequence 2, Appli
7	352	8.6	1972	9	US-10-171-311-162	Sequence 162, App
8	352	8.6	1979	9	US-09-927-597-4	Sequence 4, Appli
9	333	8.1	1979	9	US-10-205-823-419	Sequence 419, App
10	326	8.0	2383	9	US-10-082-830-260	Sequence 260, App
11	319	7.8	868	9	US-09-884-001-19	Sequence 19, Appl
12	316.5	7.7	1203	9	US-10-097-340-43	Sequence 43, Appl
13	298	7.3	751	10	US-09-864-761-38419	Sequence 38419, A
14	294	7.2	1388	9	US-10-146-473-82	Sequence 82, Appl
15	292.5	7.1	1184	10	US-09-815-242-5229	Sequence 5229, Ap
16	292.5	7.1	1188	10	US-09-815-242-12125	Sequence 12125, A
17	292.5	7.1	2310	9	US-09-991-496-120	Sequence 120, App
18	292.5	7.1	2310	9	US-09-820-843A-114	Sequence 114, App
19	292.5	7.1	2310	10	US-09-874-923-120	Sequence 120, App
20	290	7.1	2871	9	US-10-146-473-41	Sequence 41, Appl
21	287.5	7.0	2053	9	US-10-017-216-2	Sequence 2, Appli
22	285.5	7.0	2354	9	US-09-820-843A-113	Sequence 113, App
23	285	7.0	1805	9	US-09-820-843A-73	Sequence 73, Appl
24	283.5	6.9	3899	9	US-10-171-311-4	Sequence 4, Appli
25	283.5	6.9	3907	9	US-10-171-311-2	Sequence 2, Appli
26	283.5	6.9	3917	9	US-10-171-311-8	Sequence 8, Appli
27	283.5	6.9	3925	9	US-10-171-311-6	Sequence 6, Appli
28	272	6.6	923	9	US-09-820-843A-112	Sequence 112, App
29	270	6.6	660	10	US-09-864-761-47959	Sequence 47959, A
30	269.5	6.6	1597	9	US-10-017-216-6	Sequence 6, Appli
31	269.5	6.6	2055	9	US-10-017-216-4	Sequence 4, Appli
32	268.5	6.6	1958	12	US-10-028-946-4	Sequence 4, Appli
33	268.5	6.6	2054	12	US-10-028-946-2	Sequence 2, Appli
34	267	6.5	2139	9	US-10-023-219-4	Sequence 4, Appli
35	267	6.5	2139	10	US-09-727-384-6	Sequence 6, Appli
36	263	6.4	1192	10	US-09-815-242-10903	Sequence 10903, A
37	262.5	6.4	1179	10	US-09-815-242-13608	Sequence 13608, A
38	261.5	6.4	888	9	US-09-893-519A-73	Sequence 73, Appl
39	261	6.4	1045	10	US-09-815-242-10617	Sequence 10617, A
40	260.5	6.4	1641	9	US-10-017-216-5	Sequence 5, Appli
41	259	6.3	879	9	US-10-108-605-217	Sequence 217, App
42	253.5	6.2	1033	9	US-09-820-843A-75	Sequence 75, Appl
43	253.5	6.2	2125	10	US-09-919-172-29	Sequence 29, Appl
44	252.5	6.2	1179	10	US-09-815-242-13262	Sequence 13262, A
45	248	6.1	2568	10	US-09-866-108-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-884-001-2
; Sequence 2, Application US/09884001
; Publication No. US20020182656A1
; GENERAL INFORMATION:
; APPLICANT: Bird, Timothy A.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Willis, Cynthia R.
; TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
; TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)
; FILE REFERENCE: Immunex GNK/sGNK PCT
; CURRENT APPLICATION NUMBER: US/09/884,001
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/113,003
; PRIOR FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-001-2

Query Match 100.0%; Score 4095; DB 9; Length 824;
Best Local Similarity 100.0%; Pred. No. 4.8e-187;
Matches 824; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ 60

Qy         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLE 120

Qy        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
```

```
Db        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Qy        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300
Db        241 EEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALV 300

Qy        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360
Db        301 NGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKA 360

Qy        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420
Db        361 GLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHE 420

Qy        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480
Db        421 DGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQA 480

Qy        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
Db        481 LTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600
Db        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADG 600

Qy        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
Db        601 GTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
Db        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
Db        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
Db        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
```

RESULT 2
US-09-884-001-18
; Sequence 18, Application US/09884001
; Publication No. US20020182656A1
; GENERAL INFORMATION:
; APPLICANT: Bird, Timothy A.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Willis, Cynthia R.
; TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
; TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)
; FILE REFERENCE: Immunex GNK/sGNK PCT
; CURRENT APPLICATION NUMBER: US/09/884,001
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/113,003
; PRIOR FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-001-18

Query Match 61.9%; Score 2536; DB 9; Length 821;
Best Local Similarity 64.0%; Pred. No. 4.4e-113;
Matches 527; Conservative 110; Mismatches 149; Indels 38; Gaps 8;

```
Qy         23 RAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEAF 82
Db         14 KTEIERLTKELTETTHEKIQAAEYGLVVLEEKLTLKQQYDELEAEYDSLKQELEQLKEAF 73

Qy         83 GQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERL 142
Db         74 GQSFSIHRKVAEDGETREETLLQESASKEAYYLGKILEMQNELKQSRAVVTNVQAENERL 133

Qy        143 ASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQ 202
Db        134 TAVVQDLKENNEMVELQRIRMKDEIREYKFREARLLQDYTELEEENITLQKLVSTLKQNQ 193

Qy        203 VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELS 262
Db        194 VEYEGLKHEIKRFEEETVLLNSQLEDAIRLKEIAEHQLEEALETLKNEREQKNNLRKELS 253

Qy        263 HYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKK 322
Db        254 QYISLND----NHISISVDGLKFAEDGSEPNNDDK--MNGHIHGPLVKLNGDYRTPTLRK 307

Qy        323 -EGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
Db        308 GESLNP-----VSDLFSELNISEIQKLKQQLMQVEREKAILLANLQESQTQLEHTKGALT 362

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
Db        363 EQHERVHRLTEHVNAMRGLQSSKELKAELDGEKGRDSGEEAHDYEVDINGLEILECKYRV 422

Qy        442 AVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLAR 501
Db        423 AVTEVIDLKAEIKALKEKYNKSVENYTDEKAKYESKIQMYDEQVTSLEKTTKESGEKMAH 482

Qy        502 LEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQ 561
Db        483 MEKELQKMTSIANENHSTLNTAQDELVTFSEELAQLYHHVCLCNNETPNRVMLDYYRQSR 542

Qy        562 GGAGRTSPGGRTSP-EARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSP---- 616
Db        543 ----VTRSGSLKGPDDPRGLLSPRLARRGVSSPVETRTSSEPVAKESTEPSKEPSPTKTP 598

Qy        617 ----------------LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
Db        599 TISPVITAPPSSPVLDTSDIRKEPMNIYNLNAIIRDQIKHLQKAVDRSLQLSRQRAAARE 658

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
Db        659 LAPMIDKDKEALMEEILKLKSLLSTKREQIATLRAVLKANKQTAEVALANLKNKYENEKA 718

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
Db        719 MVTETMTKLRNELKALKEDAATFSSLRTMFATRCDEYVTQLDEMQRQLAAAEDEKKTLNT 778

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
Db        779 LLRMAIQQKLALTQRLEDLEFDHEQSRRSKGKLG-KSKIGSPKV 821
```

RESULT 3
US-10-097-340-31
; Sequence 31, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS

```
;  APPLICANT:  Robert C. BAST, Jr.
;  APPLICANT:  Karen LU
;  APPLICANT:  Rosemarie SCHMANDT
;  APPLICANT:  Xumei ZHAO
;  APPLICANT:  Karen GLATT
;  TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;  TITLE OF INVENTION:  Assessment, Prevention, and Therapy of Ovarian Cancer
;  FILE REFERENCE: MRI-030
;  CURRENT APPLICATION NUMBER: US/10/097,340
;  CURRENT FILING DATE:  2002-03-14
;  PRIOR APPLICATION NUMBER: 60/276,025
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/325,149
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/276,026
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/324,967
;  PRIOR FILING DATE: 2001/09/26
;  PRIOR APPLICATION NUMBER: 60/311,732
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: 60/325,102
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/323,580
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 975
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-097-340-31

  Query Match             61.9%;  Score 2536;  DB 9;  Length 975;
  Best Local Similarity   64.0%;  Pred. No. 5.3e-113;
  Matches  527;  Conservative  110;  Mismatches  149;  Indels   38;  Gaps    8;

Qy         23 RAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEAF 82
              : |::||: || ||| |||||||||| ||||| || |::||| :|:::: |:||||||
Db         14 KTEIERLTKELTETTHEKIQAAEYGLVVLEEKLTLKQQYDELEAEYDSLKQELEQLKEAF 73

Qy         83 GQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERL 142
              ||: : |:||| |||:|||:|:|||||| ||: |:||:| |||| | |:|| |:|||||
Db         74 GQSFSIHRKVAEDGETREETLLQESASKEAYYLGKILEMQNELKQSRAVVTNVQAENERL 133

Qy        143 ASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQ 202
               :| |:||| |: ||:|| |::|:|:|||||||||||||:|||||||:||| || |:|||
Db        134 TAVVQDLKENNEMVELQRIRMKDEIREYKFREARLLQDYTELEEENITLQKLVSTLKQNQ 193

Qy        203 VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELS 262
              ||:||||||||| |||| ||||||||||||||:| |||||||||| |||||||:||||||
Db        194 VEYEGLKHEIKRFEEETVLLNSQLEDAIRLKEIAEHQLEEALETLKNEREQKNNLRKELS 253

Qy        263 HYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKK 322
               |:|:||    :|: :|:|||||::| :||||| :  :||  || | ||  | :| | :|
Db        254 QYISLND----NHISISVDGLKFAEDGSEPNNDDK--MNGHIHGPLVKLNGDYRTPTLRK 307

Qy        323 -EGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
               | | |     |||| |||||||||||||||||:||||| ||| ||::| |||||:|:|:
Db        308 GESLNP-----VSDLFSELNISEIQKLKQQLMQVEREKAILLANLQESQTQLEHTKGALT 362

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
              || |:| ||||:::|:| ||:||| :  || || ||| |:   |||||| ||| ||| |
Db        363 EQHERVHRLTEHVNAMRGLQSSKELKAELDGEKGRDSGEEAHDYEVDINGLEILECKYRV 422

Qy        442 AVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLAR 501
               || |  :|: ::|||:  :      : :|| :|||:: |   |:|: ||| :::  | :|
Db        423 AVTEVIDLKAEIKALKEKYNKSVENYTDEKAKYESKIQMYDEQVTSLEKTTKESGEKMAH 482

Qy        502 LEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQ 561
              :||||:|:: :| |   :|: |||||||||||| |||||:||||||||||||||: :
Db        483 MEKELQKMTSIANENHSTLNTAQDELVTFSEELAQLYHHVCLCNNETPNRVMLDYYRQSR 542

Qy        562 GGAGRTSPGGRTSP-EARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSP---- 616
```

```
                  | |    | : ||  || |  :|: :|   |          :  |  |||
Db        543 ----VTRSGSLKGPDDPRGLLSPRLARRGVSSPVETRTSSEPVAKESTEPSKEPSPTKTP 598

Qy        617 ---------------LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQE 660
                              || |:|||||||| |||||||||| |||||: :||||| |::|
Db        599 TISPVITAPPSSPVLDTSDIRKEPMNIYNLNAIIRDQIKHLQKAVDRSLQLSRQRAAARE 658

Qy        661 LGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKA 720
              | | :|||||||||||||||||||||||||||| ||| |||||||||||||||:|||||||
Db        659 LAPMIDKDKEALMEEILKLKSLLSTKREQIATLRAVLKANKQTAEVALANLKNKYENEKA 718

Qy        721 MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNS 780
              |||||| |||||||||||||||||||| |||||||||:|||||||||||||||||||||:
Db        719 MVTETMTKLRNELKALKEDAATFSSLRTMFATRCDEYVTQLDEMQRQLAAAEDEKKTLNT 778

Qy        781 LLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTKPATPSL 824
              |||||||||||||||| || ||||:|| : |   |:|  :| :
Db        779 LLRMAIQQKLALTQRLEDLEFDHEQSRRSKGKLG-KSKIGSPKV 821


RESULT 4
US-09-746-783-16
; Sequence 16, Application US/09746783
; Publication No. US20030044935A1
;  GENERAL INFORMATION:
;       APPLICANT: Jacobs, Kenneth
;                  McCoy, John M.
;                  LaVallie, Edward R.
;                  Racie, Lisa A.
;                  Treacy, Maurice
;                  Spaulding, Vikki
;                  Agostino, Michael J.
;                  Howes, Steven H.
;                  Fechtel, Kim
;       TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;                           ENCODING THEM
;       NUMBER OF SEQUENCES: 231
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genetics Institute, Inc.
;            STREET: 87 CambridgePark Drive
;            CITY: Cambridge
;            STATE: MA
;            COUNTRY: U.S.A.
;            ZIP: 02140
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/746,783
;            FILING DATE: 21-Dec-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Milasincic, Debra J.
;            REGISTRATION NUMBER: 46,931
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 227-7400
;            TELEFAX: (617) 742-4214
;  INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 170 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-746-783-16

  Query Match             10.3%;  Score 420;  DB 9;  Length 170;
  Best Local Similarity   49.7%;  Pred. No. 2.1e-13;
  Matches   84;  Conservative   38;  Mismatches   47;  Indels    0;  Gaps    0;
```

```
Qy        353 MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDN 412
              ||:||||| ||| ||:::| ||||||:|:|:|| |:| ||||:::|:| ||:||| :  ||
Db          1 MQVEREKAILLANLQESQTQLEHTKGALTEQHERVHRLTEHVNAMRGLQSSKELKAELDG 60

Qy        413 EKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKG 472
              || ||| |:   ||||||| ||| ||| ||| |  :|: ::|||:  :      : :||
Db         61 EKGRDSGEEAHDYEVDINGLEILECKYRVAVTEVIDLKAEIKALKEKYNKSVENYTDEKA 120

Qy        473 RYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLS 521
              :||:: |   |:|: ||| :::  | :| :|||||:|:: :| |   :|:
Db        121 KYESKIQMYDEQVTSLEKTTKESGEKMAHMEKELQKMTSIANENHSTLN 169


RESULT 5
US-10-171-311-164
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164
;   LENGTH: 1938
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-164

  Query Match             8.6%;  Score 352;  DB 9;  Length 1938;
  Best Local Similarity   20.9%;  Pred. No. 4.4e-09;
  Matches  187;  Conservative  176;  Mismatches  342;  Indels  188;  Gaps   31;

Qy          9 EYARLVMEAQPEWLRAEVKRLSHEL------AETTREKIQAAEYGLAVLEEKHQLKLQFE 62
              :: ||  : :|     :| |   |:       : |:|: | ||  |  ||:||
Db        833 QWWRLFTKVKP---LLQVTRQEEEMQAKEDELQKTKERQQKAENELKELEQKH------S 883

Qy         63 ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQ 122
              :|  :   :: :::   | : :|     ::||  :  || | :  |  |:   |   :||
Db        884 QLTEEKNLLQEQLQAETELYAEAEEMRVRLAAKKQELEEILHEMEARLEEEEDRG-QQLQ 942

Qy        123 TELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD-- 180
               | |::   : : : : |   :  |:|:       | :  :| |:|         |: |
Db        943 AERKKMAQQMLDLEEQLEEEEAARQKLQLEKVTAEAKIKKLEDEI---------LVMDDQ 993

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ::| :|   |::::| |  |  | |     : :|: : | : |:||  :: :| | ::|
Db        994 NNKLSKERKLLEERISDLTTNLAEEEEKAKNLTKLKNKHESMISELEVRLKKEEKSRQEL 1053

Qy        241 EEALETLKTE----REQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN-- 294
              |:    |: :     ||   |: ::: : :   :     | :|   :  |: |: ||
Db       1054 EKLKRKLEGDASDFHEQIADLQAQIAE-LKMQLAKKEEELQAAL--ARLDDEIAQKNNAL 1110

Qy        295 ----DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQ 350
                  : |  :: :    ::    ||    |:: |     :| ::|   |: : |   |
Db       1111 KKIRELEGHISDLQEDLDSERAARNKAEKQKRD-LGEELEALKTELEDTLDSTATQ---Q 1166
```

```
Qy        351 QLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQE-------KVTRLTENLSALRRLQAS 403
              :|      ::  :|     |      | ||   :: ||         |  ||| |   :| :|:
Db       1167 ELRAKREQEVTVLKKALD-----EETRSHEAQVQEMRQKHAQAVEELTEQLEQFKRAKAN 1221

Qy        404 KERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAR 463
                 ||  |      |: |                          |||||   :|       :
Db       1222 ------LDKNKQTLEKENADL--------------------AGELRVLGQA------KQ 1248

Qy        464 EAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVA 523
              | :|  :| : ||: | |  | |  |:|  :  : : :|: |:: |: :  | :|
Db       1249 EVEH--KKKKLEAQVQELQSKCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKL 1306

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
              :: : | :| :      :   ||  :: :   :  |    | |   :   |   :::
Db       1307 AKDVASLSSQLQDTQE---LLQEETRQKLNVS-TKLRQLEEERNSLQDQLDEEMEAKQN- 1361

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQA 643
              | : :                      |:|   ||| :::  :  : :  : :  | |
Db       1362 --LERHI-------------------STLNIQLSDSKKKLQDFASTVEALEEGKKRFQK 1399

Qy        644 AVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTL-RTVLKANKQ 702
              ::  |:   :: |: :    ::| |  | :|:  |   |   :|: :: | :   | ::
Db       1400 EIENLTQQYEEKAAAYD---KLEKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQRKFDQL 1456

Qy        703 TAEVALANLKSKYENEK----------------AMVTETMMKLRNE-------LKALKE 738
              ||    |: ||| :|:                  |   |  :: : |       |||  |
Db       1457 LAEE--KNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAEME 1514

Qy        739 DAATF-----------SSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
              |  :              | ||:  |: :|  |||:|:: :| | || |      ||: :
Db       1515 DLVSSKDDVGKNVHELEKSKRAL-ETQMEEMKTQLEELEDELQATEDAK------LRLEV 1567

Qy        787 --------------------QQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
                                 :::  | ::|   | : |  |: || ||   |
Db       1568 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKK 1620


RESULT 6
US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
;  APPLICANT: Malik, Fady
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Craven, Andrew
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Hartman, James
;  TITLE OF INVENTION: Human smooth muscle myosin heavy chain
;  FILE REFERENCE: CYTOP018
;  CURRENT APPLICATION NUMBER: US/09/927,597
;  CURRENT FILING DATE:  2001-08-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1945
;   TYPE: PRT
;   ORGANISM: Human
US-09-927-597-2

  Query Match             8.6%;  Score 352;  DB 9;  Length 1945;
  Best Local Similarity   20.9%;  Pred. No. 4.4e-09;
  Matches  187;  Conservative  176;  Mismatches  342;  Indels  188;  Gaps   31;

Qy          9 EYARLVMEAQPEWLRAEVKRLSHEL------AETTREKIQAAEYGLAVLEEKHQLKLQFE 62
              :: ||  : :|     :| |   |:       : |:|: | ||  |  ||:||
Db        840 QWWRLFTKVKP---LLQVTRQEEEMQAKEDELQKTKERQQKAENELKELEQKH------S 890

Qy         63 ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQ 122
              :|  :   :: :::   | : :|     ::||  :  || | :  |  |:   |   :||
Db        891 QLTEEKNLLQEQLQAETELYAEAEEMRVRLAAKKQELEEILHEMEARLEEEEDRG-QQLQ 949
```

```
Qy        123 TELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD-- 180
Db        950 AERKKMAQQMLDDLEEQLEEEEAARQKLQLEKVTAEAKIKKLEDEI---------LVMDDQ 1000

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
Db        1001 NNKLSKERKLLEERISDLTTNLAEEEEKAKNLTKLKNKHESMISELEVRLKKEEKSRQEL 1060

Qy        241 EEALETLKTE----REQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN-- 294
Db        1061 EKLKRKLEGDASDFHEQIADLQAQIAE-LKMQLAKKEEELQAAL--ARLDDEIAQKNNAL 1117

Qy        295 ----DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQ 350
Db        1118 KKIRELEGHISDLQEDLDSERAARNKAEKQKRD-LGEELEALKTELEDTLDSTATQ---Q 1173

Qy        351 QLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQE-------KVTRLTENLSALRRLQAS 403
Db        1174 ELRAKREQEVTVLKKALD-----EETRSHEAQVQEMRQKHAQAVEELTEQLEQFKRAKAN 1228

Qy        404 KERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAR 463
Db        1229 ------LDKNKQTLEKENADL--------------------AGELRVLGQA------KQ 1255

Qy        464 EAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVA 523
Db        1256 EVEH--KKKKLEAQVQELQSKCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKL 1313

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
Db        1314 AKDVASLSSQLQDTQE---LLQEETRQKLNVS-TKLRQLEEERNSLQDQLDEEMEAKQN- 1368

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQA 643
Db        1369 --LERHI--------------------STLNIQLSDSKKKLQDFASTVEALEEGKKRFQK 1406

Qy        644 AVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTL-RTVLKANKQ 702
Db        1407 EIENLTQQYEEKAAAYD---KLEKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQRKFDQL 1463

Qy        703 TAEVALANLKSKYENEK----------------AMVTETMMKLRNE-------LKALKE 738
Db        1464 LAEE--KNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAEME 1521

Qy        739 DAATF-----------SSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
Db        1522 DLVSSKDDVGKNVHELEKSKRAL-ETQMEEMKTQLEELEDELQATEDAK------LRLEV 1574

Qy        787 --------------------QQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
Db        1575 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKK 1627
```

```
RESULT 7
US-10-171-311-162
; Sequence 162, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-162

  Query Match             8.6%;  Score 352;  DB 9;  Length 1972;
  Best Local Similarity   20.9%;  Pred. No. 4.5e-09;
  Matches  187;  Conservative  176;  Mismatches  342;  Indels  188;  Gaps   31;

Qy          9 EYARLVMEAQPEWLRAEVKRLSHEL------AETTREKIQAAEYGLAVLEEKHQLKLQFE 62
Db        833 QWWRLFTKVKP---LLQVTRQEEEMQAKEDELQKTKERQQKAENELKELEQKH------S 883

Qy         63 ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQ 122
Db        884 QLTEEKNLLQEQLQAETELYAEAEEMRVRLAAKKQELEEILHEMEARLEEEEDRG-QQLQ 942

Qy        123 TELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD-- 180
Db        943 AERKKMAQQMLDLEEQLEEEEAARQKLQLEKVTAEAKIKKLEDEI---------LVMDDQ 993

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
Db        994 NNKLSKERKLLEERISDLTTNLAEEEEKAKNLTKLKNKHESMISELEVRLKKEEKSRQEL 1053

Qy        241 EEALETLKTE----REQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN-- 294
Db        1054 EKLKRKLEGDASDFHEQIADLQAQIAE-LKMQLAKKEEELQAAL--ARLDDEIAQKNNAL 1110

Qy        295 ----DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQ 350
Db        1111 KKIRELEGHISDLQEDLDSERAARNKAEKQKRD-LGEELEALKTELEDTLDSTATQ---Q 1166

Qy        351 QLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQE-------KVTRLTENLSALRRLQAS 403
Db        1167 ELRAKREQEVTVLKKALD-----EETRSHEAQVQEMRQKHAQAVEELTEQLEQFKRAKAN 1221

Qy        404 KERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAR 463
Db        1222 ------LDKNKQTLEKENADL--------------------AGELRVLGQA------KQ 1248

Qy        464 EAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVA 523
Db        1249 EVEH--KKKKLEAQVQELQSKCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKL 1306

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
Db        1307 AKDVASLSSQLQDTQE---LLQEETRQKLNVS-TKLRQLEEERNSLQDQLDEEMEAKQN- 1361

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQA 643
Db        1362 --LERHI--------------------STLNIQLSDSKKKLQDFASTVEALEEGKKRFQK 1399

Qy        644 AVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTL-RTVLKANKQ 702
Db        1400 EIENLTQQYEEKAAAYD---KLEKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQRKFDQL 1456

Qy        703 TAEVALANLKSKYENEK----------------AMVTETMMKLRNE-------LKALKE 738
Db        1457 LAEE--KNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAEME 1514

Qy        739 DAATF-----------SSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
Db        1515 DLVSSKDDVGKNVHELEKSKRAL-ETQMEEMKTQLEELEDELQATEDAK------LRLEV 1567
```

```
Qy        787 --------------------QQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
Db       1568 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKK 1620
```

RESULT 8
US-09-927-597-4
; Sequence 4, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-4

Query Match 8.6%; Score 352; DB 9; Length 1979;
Best Local Similarity 20.9%; Pred. No. 4.5e-09;
Matches 187; Conservative 176; Mismatches 342; Indels 188; Gaps 31;

```
Qy          9 EYARLVMEAQPEWLRAEVKRLSHEL------AETTREKIQAAEYGLAVLEEKHQLKLQFE 62
Db        840 QWWRLFTKVKP---LLQVTRQEEEMQAKEDELQKTKERQQKAENELKELEQKH------S 890

Qy         63 ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQ 122
Db        891 QLTEEKNLLQEQLQAETELYAEAEEMRVRLAAKKQELEEILHEMEARLEEEEDRG-QQLQ 949

Qy        123 TELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD-- 180
Db        950 AERKKMAQQMLDLEEQLEEEEAARQKLQLEKVTAEAKIKKLEDEI---------LVMDDQ 1000

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
Db       1001 NNKLSKERKLLEERISDLTTNLAEEEEKAKNLTKLKNKHESMISELEVRLKKEEKSRQEL 1060

Qy        241 EEALETLKTE----REQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN-- 294
Db       1061 EKLKRKLEGDASDFHEQIADLQAQIAE-LKMQLAKKEEELQAAL--ARLDDEIAQKNNAL 1117

Qy        295 ----DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQ 350
Db       1118 KKIRELEGHISDLQEDLDSERAARNKAEKQKRD-LGEELEALKTELEDTLDSTATQ---Q 1173

Qy        351 QLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQE-------KVTRLTENLSALRRLQAS 403
Db       1174 ELRAKREQEVTVLKKALD-----EETRSHEAQVQEMRQKHAQAVEELTEQLEQFKRAKAN 1228

Qy        404 KERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAR 463
Db       1229 ------LDKNKQTLEKENADL--------------------AGELRVLGQA------KQ 1255

Qy        464 EAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVA 523
Db       1256 EVEH--KKKKLEAQVQELQSKCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKL 1313

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
Db       1314 AKDVASLSSQLQDTQE--LLQEETRQKLNVS-TKLRQLEEERNSLQDQLDEEMEAKQN- 1368

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQA 643
Db       1369 --LERHI-------------------STLNIQLSDSKKKLQDFASTVEALEEGKKRFQK 1406

Qy        644 AVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTL-RTVLKANKQ 702
Db       1407 EIENLTQQYEEKAAAYD---KLEKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQRKFDQL 1463

Qy        703 TAEVALANLKSKYENEK----------------AMVTETMMKLRNE-------LKALKE 738
Db       1464 LAEE--KNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAEME 1521

Qy        739 DAATF-----------SSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
Db       1522 DLVSSKDDVGKNVHELEKSKRAL-ETQMEEMKTQLEELEDELQATEDAK------LRLEV 1574

Qy        787 --------------------QQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
Db       1575 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKK 1627
```

RESULT 9
US-10-205-823-419
; Sequence 419, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-419

Query Match 8.1%; Score 333; DB 9; Length 1979;
Best Local Similarity 21.7%; Pred. No. 3.6e-08;
Matches 205; Conservative 174; Mismatches 354; Indels 212; Gaps 41;

```
Qy         15 MEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAV--LEEKHQLKL-QFEELEVDYEAI 71
Db        721 IEAELCWAKKRLLEEANKYEKTIEELSNARNLNTSALQLEHEHLIKLNQKKDMEI----- 775

Qy         72 RSEMEQLKEAFGQAHTNHK--KVAADGESREESLIQESASKEQYYVRKVLELQTELKQLR 129
Db        776 ----AELKKNIEQMDTDHKETKDVLSSSLEEQKQLTQLINKKEIFIEKLKERSSKLQEEL 831

Qy        130 NVLTNTQSENERLASVAQE-------LKEINQNVEIQRGRLR--------------DDI 167
```

```
Db        832 DKYSQALRKNEILRQTIEEKDRSLGSMKEENNHLQEELERLREEQSRTAPVADPKTLDSV 891

Qy        168 KEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVE-FEGLKHEIKRLEE---ETEYLN 223

Db        892 TELASEVSQLNTIKEHLEEEIKHHQKIIEDQNQSKMQLLQSLQEQKKEMDEFRYQHEQMN 951

Qy        224 SQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGL 283

Db        952 ATHTQLFLEKDEEIKSLQKTIEQIKTQ------LHEERQDIQTDNSDIFQETKVQSLN-- 1003

Qy        284 KFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNIS 343

Db       1004 ---------------IENGSEKHDLSKAETERLVKGIKER-------ELEIKLLNEKNIS 1041

Qy        344 EIQKLKQQLMQMEREKAGLLATL---QDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRL 400

Db       1042 ----LTKQIDQLSKDEVGKLTQIIQQKDLEIQALHARISSTSHTQDVVYLQQQLQA---Y 1094

Qy        401 QASKERQTALDNEKDRD-SHEDGDYYE-VDINGPEILACKYHVAVAEAG--ELREQLKAL 456

Db       1095 AMEREKVFAVLNEKTRENSHLKTEYHKMMDI-----------VAAKEAALIKLQDENKKL 1143

Qy        457 RSTHEA------RE-----AQHAEEKGRYEAEGQALTEK----VSLLEKA---------- 491

Db       1144 STRFESSGQDMFRETIQNLSRIIREK---DIEIDALSQKCQTLLAVLQTSSTGNEAGGVN 1200

Qy        492 SRQDRELLA---RLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNET 548

Db       1201 SHQFEELLQERDKLKQQVKKMEEWKQQVMTTVQNMQHESAQLQEELHQLQAQV-LVDSDN 1259

Qy        549 PNRVMLDY------YREG-------------QGGAGRTSPGGRTSPEARGRR---SPIL 585

Db       1260 NSKLQVDYTGLIQSYEQNETKLKNFGQELAQVQHSIGQLC---NTKDLLLGKLDIISPQL 1316

Qy        586 LPKGLLAPEAGRADGGTGDSSPSPGSS-LPSPLSDPRREPMNIYNLIAIIRDQIKHLQAA 644

Db       1317 SSASLLTPQSAECLRASKSEVLSESSELLQQELEELRKSLQEKDATIRTLQENNHRLSDS 1376

Qy        645 VDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL-STKREQITTLRT--VLKANK 701

Db       1377 IAATSELERKEHEQTDSEIKQLKEKQDVLQKLLKEKDLLIKAKSDQLLSSNENFTNKVNE 1436

Qy        702 -QTAEVALANLKS------------KYENEKAMVT------------ETMMK----LRN- 731

Db       1437 NELLRQAVTNLKERILILEMDIGKLKGENEKIVETYRGKETEYQALQETNMKFSMMLREK 1496

Qy        732 --ELKALKEDAATFSSL-------------RAMFATRCDEYITQLDEMQR-QLAAAEDEK 775

Db       1497 EFECHSMKEKALAFEQLLKEKEQGKTGELNQLLNAVKSMQEKTVVFQQERDQVMLALKQK 1556

Qy        776 KTLNSLLRMAIQ----QKLALTQRLE-----LLELDHEQTRRGRA 811

Db       1557 QMENTALQNEVQRLRDKEFRSNQELERLRNHLLESEDSYTREALA 1601


RESULT 10
US-10-082-830-260
; Sequence 260, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
;  APPLICANT: Sun, Yongming
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Salceda, Susana
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Turner, Leah
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
;  TITLE OF INVENTION:  Genes and Proteins
;  FILE REFERENCE: DEX-0249
;  CURRENT APPLICATION NUMBER: US/10/082,830
;  CURRENT FILING DATE:  2001-10-29
;  PRIOR APPLICATION NUMBER: 60/243,802
;  PRIOR FILING DATE: 2000-10-27
;  NUMBER OF SEQ ID NOS: 282
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 260
;   LENGTH: 2383
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-082-830-260

  Query Match             8.0%;  Score 326;  DB 9;  Length 2383;
  Best Local Similarity   21.8%;  Pred. No. 9.4e-08;
  Matches  229;  Conservative  145;  Mismatches  378;  Indels  300;  Gaps   42;

Qy          5 SEEEEY---ARLVMEAQPEWLRAE---VKRLSHEL------AETTREKIQAAEYGLAVLE 52

Db        437 SKERELLQKAREELRQQLEVLEQEAWRLRRVNVELQLQGDSAQGQKEE-QQEELHLAV-R 494

Qy         53 EKHQLKLQFEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGE------------SR 99

Db        495 ERERLQEMLMGLEAKQSESLSELITLREALESSH-------LEGELLRQEQTEVTAALAR 547

Qy        100 EESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERLA-------SVAQELKEI 152

Db        548 AEQSIAELSSSEN-------TLKTEVADLRAAAVKLSALNEALALDKVGLNQQLLQLEEE 600

Qy        153 NQNV-------EIQRGRLRDDIKEYKFRE---------------------ARLLQDYSEL 184

Db        601 NQSVCSRMEAAEQARNALQVDLAEAEKRREALWEKNTHLEAQLQKAEEAGAELQADLRDI 660

Qy        185 EEENISLQKQVSVLRQNQ----VEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240

Db        661 QEEKEEIQKKLSESRHQQEAATTQLEQLHQEAKRQEEVLARAVQEKEALVREKAALEVRL 720

Qy        241 E--------------------EALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSL 280

Db        721 QAVERDRQDLAEQLQGLSSAKELLESSLFEAQQQNSVIEVTKGQLEVQIQTVTQAKEV-I 779

Qy        281 DG----LKFSDDA----AEPNNDAEA--LVNGFEHGGLAKLPLDNKTSTPKKE------- 323

Db        780 QGEVRCLKLELDTERSQAEQERDAAARQLAQAEQEG---KTALEQQKAAHEKEVNQLREK 836

Qy        324 ----------GLAPPSPSLVSDLLS-ELNISEIQ------------------KLKQQLM 353

Db        837 WEKERSWHQQELAKALESLEREKMELEMRLKEQQTEMEAIQAQREEERTQAESALCQMQL 896

Qy        354 QMEREKAGLLATLQDTQKQL--------------------EHTRGSLSEQQEKVTR--- 389

Db        897 ETEKERVSLLETLLQTQKELADASQQLERLRQDMKVQKLKEQETTGILQTQLQEAQRELK 956

Qy        390 ------------LTENLSAL---------------RRLQASKERQTALDNEKDRDSHEDG 422

Db        957 EAARQHRDDLAALQEESSSLLQDKMDLQKQVEDLKSQLVAQDDSQRLVEQEVQEKLRETQ 1016

Qy        423 DYYEV----------------------------DINGPEILACKYHV--AVAEAGELRE 451

Db       1017 EYNRIQKELEREKASLTLSLMEKEQRLLVLQEADSIRQQELSALRQDMQEAQGEQKELSA 1076

Qy        452 QLKALRSTHEAREAQHAEEKGRY--EAEGQALTE---KVSLLEKASRQDRELLARLEKEL 506

Db       1077 QMELLRQEVKEKEADFLAQEAQLLEELEASHITEQQLRASLWAQEAKA-AQLQLRLRSTE 1135

Qy        507 KKVSDVAGETQ-GSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565

Db       1136 SQLEALAAEQQPGNQAQAQAQLASLYSALQQALGSVCESRPEL------------SGGG 1182

Qy        566 RTSPG-GRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREP 624

Db       1183 DSAPSVWGLEPDQNGARS--LFKRGPLL------------TALSAEAVASALHKLHQDL 1227

Qy        625 MNIYNLIAIIRDQIKHLQAAVDRT----------TELSRQRIASQELGPAVDKDKEALM 673

Db       1228 WKTQQTRDVLRDQVQKLEERLTDTEAEKSQVHTELQDLQRQLSQNQEEKSKWEGKQNSLE 1287

Qy        674 EEILKLKSLLSTKREQITTL---RTVLKANKQTAEVALANLKSKYENEKAMVTETMMKLR 730
```

```
                |:::|   ::: : ::      |   :  ::  : |  || :: |: :| | |     |
Db       1288 SELMELHETMASLQSRLRRAELQRMEAQGERELLQAAKENLTAQVEHLQAAVVEA----R 1343

Qy        731 NELKA---LKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQ 787
              :  |   |:|| |    |:    : :|      | :|: | |  |:  |      |:|
Db       1344 AQASAAGILEEDLRT---ARSALKLKNEEV-----ESERERAQALQEQGELKVAQGKALQ 1395

Qy        788 QKLA-LTQRLELLELDHEQTRRGRAKAAPKTK 818
              : || ||| |   | :  :| ||: :   | :
Db       1396 ENLALLTQTLAERE-EEVETLRGQIQELEKQR 1426
                •

RESULT 11
US-09-884-001-19
; Sequence 19, Application US/09884001
; Publication No. US20020182656A1
; GENERAL INFORMATION:
;  APPLICANT: Bird, Timothy A.
;  APPLICANT:  Peschon, Jacques J.
;  APPLICANT:  Sims, John E.
;  APPLICANT:  Virca, G. Duke
;  APPLICANT:  Willis, Cynthia R.
;  TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
;  TITLE OF INVENTION:  Containing NEK-Like Kinase (GNK)
;  FILE REFERENCE: Immunex GNK/sGNK PCT
;  CURRENT APPLICATION NUMBER: US/09/884,001
;  CURRENT FILING DATE:  2001-06-18
;  PRIOR APPLICATION NUMBER: 60/113,003
;  PRIOR FILING DATE: 1998-12-18
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 868
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-884-001-19

  Query Match             7.8%;  Score 319;  DB 9;  Length 868;
  Best Local Similarity   22.1%;  Pred. No. 7e-08;
  Matches  191;  Conservative  148;  Mismatches  361;  Indels  164;  Gaps   32;

Qy         40 KIQAAEYGLAVL--------EEKHQLKLQFEELEVDYEAIRSEMEQLKEAFGQAHTNHKK 91
              ::| ||   | |        ||| ::: :  |     ||  :::|||     |     ::
Db          7 QLQKAEEAGAELQADLRDIQEEKEEIQKKLSESRHQQEAATTQLEQLH----QEAKRQEE 62

Qy         92 VAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKE 151
              | |     :|:|::| |: |     ::  :: : : |  |    |  | | | |   | ::
Db         63 VLARAVQEKEALVREKAALEV----RLQAVERDRQDLAEQLQGLSSAKELLESSLFEAQQ 118

Qy        152 INQNVEIQRGRLRDDI------KEYKFREARLLQ-----DYSELEEENISLQKQVSVLRQ 200
              |  :|: :|:|   |    ||    | | |:     : |: |:|  :  :|::
Db        119 QNSVIEVTKGQLEVQIQTVTQAKEVIQGEVRCLKLELDTERSQAEQERDAAARQLA---- 174

Qy        201 NQVEFEGL----------KHEIKRL----EEETEYLNSQLEDAIRLKEISERQLEEALET 246
              | | ||           : |: :|    |:|  :   :|  |:   |  : :||  |:
Db        175 -QAEQEGKTALEQQKAAHEKEVNQLREKWEKERSWHQQELAKALESLEREKMELEMRLKE 233

Qy        247 LKTEREQKNSLRKE--------LSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEA 298
              :|| |   : |:|        |       :    | |   |   |   ||::
Db        234 QQTEMEAIQAQREEERTQAESALCQMQLETEKERVSLLETLLQTQKELADASQQ------ 287

Qy        299 LVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMERE 358
                  | :|  | |    |::       : : :   ||  :  |  :  |  : : |
Db        288 ---------LERLRQDMKVQKLKEQETTGILQTQLQEAQRELKEAARQH-RDDLAALQEE 337

Qy        359 KAGLLATLQDTQKQLEHTRGSL-----------SEQQEKVTRLTENLSALRRLQASKERQ 407
              : ||    | |||:|  :  |            | |||:    |      |:|   ||:
Db        338 SSSLLQDKMDLQKQVEDLKSQLVAQDDSQRLVEQEVQEKLRETQE----YNRIQKELERE 393

Qy        408 TA------LDNEKDRDSHEDGDYYEVDINGPEILACKYHV--AVAEAGELREQLKALRST 459
              |      :: |:     :: |     |   |: | :  :  |  |  ||  |:: ||
Db        394 KASLTLSLMEKEQRLLVLQEAD----SIRQQELSALRQDMQEAQGEQKELSAQMELLRQE 449

Qy        460 HEAREAQHAEEKGRY--EAEGQALTE---KVSLLEKASRQDRELLARLEKELKKVSDVAG 514
              : :||    :: :   | |   :||   : ||  : ::   :|  ||      ::  :|
Db        450 VKEKEADFLAQEAQLLEELEASHITEQQLRASLWAQEAKA-AQLQLRLRSTESQLEALAA 508

Qy        515 ETQ-GSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPG-GR 572
              | | |: : || :| :    |      ||    |                  | | ::|
Db        509 EQQPGNQAQAQAQLASLYSALQQALGSVCESRPEL------------SGGGDSAPSVWG 555

Qy        573 TSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIA 632
                |:  | ||  |  :| |               :     :: | |    ::
Db        556 LEPDQNGARS--LFKRGPLL------------TALSAEAVASALHKLHQDLWKTQQTRD 600

Qy        633 IIRDQIKHLQAAVDRT----------TELSRQRIASQELGPAVDKDKEALMEEILKLKS 681
              .::|||:: |:  :  |            :| ||  .:||     :  : :|  |:::|
Db        601 VLRDQVQKLEERLTDTEAEKSQVHTELQDLQRQLSQNQEEKSKWEGKQNSLESELMELHE 660

Qy        682 LLSTKREQITTL---RTVLKANKQTAEVALANLKSKYENEKAMVTETMMKLRNELKA--- 735
              ::: : ::      |   :  ::  : |  || :: |: :| | |     | :  |
Db        661 TMASLQSRLRRAELQRMEAQGERELLQAAKENLTAQVEHLQAAVVEA----RAQASAAGI 716

Qy        736 LKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLA-LTQ 794
              |:||  |    |:    : :|      | :|: | |  |:  |      |:|: || |||
Db        717 LEEDLRT---ARSALKLKNEEV-----ESERERAQALQEQGELKVAQGKALQENLALLTQ 768

Qy        795 RLELLELDHEQTRRGRAKAAPKTK 818
              |   | :  :| ||: :   | :
Db        769 TLAERE-EEVETLRGQIQELEKQR 791


RESULT 12
US-10-097-340-43
; Sequence 43, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
;  APPLICANT: John MONAHAN
;  APPLICANT:  Manjula GANNAVARAPU
;  APPLICANT:  Sebastian HOERSCH
;  APPLICANT:  Shubhangi KAMATKAR
;  APPLICANT:  Steve G. KOVATS
;  APPLICANT:  Rachel E. MEYERS
;  APPLICANT:  Michael MORRISEY
;  APPLICANT:  Peter OLANDT
;  APPLICANT:  Ami SEN
;  APPLICANT:  Peter VEIBY
;  APPLICANT:  Gordon B. MILLS
;  APPLICANT:  Robert C. BAST, Jr.
;  APPLICANT:  Karen LU
;  APPLICANT:  Rosemarie SCHMANDT
;  APPLICANT:  Xumei ZHAO
;  APPLICANT:  Karen GLATT
;  TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
;  TITLE OF INVENTION:  Assessment, Prevention, and Therapy of Ovarian Cancer
;  FILE REFERENCE: MRI-030
;  CURRENT APPLICATION NUMBER: US/10/097,340
;  CURRENT FILING DATE:  2002-03-14
;  PRIOR APPLICATION NUMBER: 60/276,025
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/325,149
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/276,026
;  PRIOR FILING DATE: 2001-03-14
;  PRIOR APPLICATION NUMBER: 60/324,967
;  PRIOR FILING DATE: 2001/09/26
;  PRIOR APPLICATION NUMBER: 60/311,732
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: 60/325,102
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: 60/323,580
;  PRIOR FILING DATE: 2001-09-19
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
;   LENGTH: 1203
```

```
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-097-340-43

  Query Match             7.7%;  Score 316.5;  DB 9;  Length 1203;
  Best Local Similarity   21.9%;  Pred. No. 1.3e-07;
  Matches  209;  Conservative  148;  Mismatches  309;  Indels  289;  Gaps   39;

Qy          3 APSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGL-AVLEEKHQLKLQF 61
              | : : |  | | | | : |  |||:         |:|:: :: ||   |||| :   :
Db        357 AVAGQGELTRKVEELQRK-LDEEVKK---------RQKLEPSQVGLERQLEEKTEECSRL 406

Qy         62 EELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLEL 121
              :||    |  : | :|  :       | |::   ||     |             :|:||
Db        407 QEL---LERRKGEAQQSNKEL----QNMKRLLDQGEDLRHGL-----------ETQVMEL 448

Qy        122 QTELKQ---------------------LRNVLTNTQSENERL-----------ASVAQEL 149
              | :||                   |  ||   |   |:|                :: :|:
Db        449 QNKLKHVQGPEPAKEVLLKDLLETRELLEEVLEGKQRVEEQLRLRERELTALKGALKEEV 508

Qy        150 KEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVEFEGLK 209
             :| ||  | : : | :: :       ||:: || |    ||  :::|       ||:
Db        509 ASRDQEVEHVRQQYQRDTEQLRRSMQDATQDHAVLEAER---QKMSALVR-------GLQ 558

Qy        210 HEIKRLEEET-----------EYLNSQLEDAIRL---KEISERQLEEALETLKTEREQKN 255
              |::   |||           | | :  :: ::|   ||  | :| | :| |: | ||
Db        559 RELEETSEETGHWQSMFQKNKEDLRATKQELLQLRMEKEEMEEELGEKIEVLQRELEQAR 618

Qy        256 S----------LRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEH 305
             :          |:|||          |      |   : | : |  : | |
Db        619 ASAGDTRQVEVLKKEL---------LRTQEELKELQAERQSQEVAGRHRDRE-------- 661

Qy        306 GGLAKLPLDNKTSTPKKEGLAPPSPSLVSDL---LSELNISEIQKLKQQLMQ--MEREKA 360
                    |: : :  : |           :|    | | |: ::||  ||| |   |  ||
Db        662 -------LEKQLAVLRVE----------ADRGRELEEQNL-QLQKTLQQLRQDCEEASKA 703

Qy        361 GLLA---------------TLQDTQ-------------KQLEHTRGSLSEQQEKVTRL 390
              ::|               ||::||             :||: ||| :   :    ||
Db        704 KMVAEAEATVLGQRRAAVETTLRETQEENDEFRRRILGLEQQLKETRGLVDGGEAVEARL 763

Qy        391 TENLSAL--------RRLQASKERQTALDNEK----------------------DRDS 418
              : |  |         | ||:| : :|    |                      :|
Db        764 RDKLQRLEAEKQQLEEALNASQEEEGSLAAAKRALEARLEEAQRGLARLGQEQQTLNRAL 823

Qy        419 HEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEG 478
              |:|   ||   |                 || || : |  | : |  :  |: |  |
Db        824 EEEGKQREVLRRG-------------KAELEEQKRLLDRTVD-RLNKELEKIG--EDSK 866

Qy        479 QALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLY 538
              ||| :  :.||    : |  :|  ::: |  :  | :| | ||  |||:    : |
Db        867 QALQQLQAQLEDYKEKARREVADAQRQAKDWASEAEKTSGGLSRLQDEIQRLRQAL---- 922

Qy        539 HHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRA 598
              :|  |:  :   :  |  : :  : | .: | |:      |||
Db        923 -QASQAERDTA-RLDKELLAQRLQGLEQEAENKKRSQDDRARQL-----KGL-------- 967

Qy        599 DGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDR-TTELSRQRIA 657
                  | | : | : :       | : :: |::   :  ||:  ||| ::| |
Db        968 --------EEKVSRLETELDEEK-------NTVELLTDRVNRGRDQVDQLRTELMQERSA 1012

Qy        658 SQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSK--- 714
              |:|    : || :|  :   ||: |::                |    :|: |:|:
Db       1013 RQDL----ECDKISLERQNKDLKTRLASSE-----------GFQKPSASLSQLESQNQL 1056

Qy        715 -------YENEKAMVTETMMKLRNELKALK---EDAATFSSLRAMFATRCDEYITQLDEM 764
                  | || ::  |  ||  ::| |    ||       |     : |:   ::  :
Db       1057 LQERLQAEEREKTVLQSTNRKLERKVKELSIQIEDE------RQHVNDQKDQLSLRVKAL 1110

Qy        765 QRQLAAAEDEKKTLNSLLRMA---IQQKLALTQRLE--LLELDHEQTRRGRAKAA 814
              :||:  ||:| : |: | : |   ::::  : ::|:  :  |: :  |:    ||
Db       1111 KRQVDEAEEEIERLDGLRKKAQREVEEQHEVNEQLQARIKSLEKDSWRKASRSAA 1165
```

```
RESULT 13
US-09-864-761-38419
; Sequence 38419, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38419
;   LENGTH: 751
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL049872.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.72
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
;   OTHER INFORMATION: EST_HUMAN HIT: AU138211.1, EVALUE 2.00e-83
;   OTHER INFORMATION: SWISSPROT HIT: Q10411, EVALUE 3.00e-18
US-09-864-761-38419

  Query Match             7.3%;  Score 298;  DB 10;  Length 751;
```

```
  Best Local Similarity   21.9%;  Pred. No. 6e-07;
  Matches  170;  Conservative  145;  Mismatches  301;  Indels  160;  Gaps   32;

Qy         74 EMEQLKEAFGQAHTNHK--KVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNV 131
              |: :||:   |  |:||  |        |:  : :  :|:: :: |: |  ::|::  :
Db          5 EIAELKKNIEQMDTDHKETKDVLSSSLEEQKQLTQLINKKEIFIEKLKERSSKLQEELDK 64

Qy        132 LTNTQSENERLASVAQE-------LKEINQNVEIQRGRLR---------------DDIKE 169
               :   :|| |   :|       :|| | ::: :  |||                | : |
Db         65 YSQALRKNEILRQTIEEKDRSLGSMKEENNHLQEELERLREEQSRTAPVADPKTLDSVTE 124

Qy        170 YKFREARLLQDYSELEEENISLQKQVSVLRQNQVE-FEGLKHEIKRLEE---ETEYLNSQ 225
                ::|      ||||    || :    |::::  : |: : | ::|   : | :|:
Db        125 LASEVSQLNTIKEHLEEEIKHHQKIIEDQNQSKMQLLQSLQEQKKEMDEFRYQHEQMNAT 184

Qy        226 LEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKF 285
                  |:   : |:: :| :||:      | :|     : |   :       ||:
Db        185 HTQLFLEKDEEIKSLQKTIEQIKTQ------LHEERQDIQTDNSDIFQETKVQSLN---- 234

Qy        286 SDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEI 345
                    : || |   |:|   :      |:          |   ||:| |||
Db        235 -------------IENGSEKHDLSKAETERLVKGIKER-------ELEIKLLNEKNIS-- 272

Qy        346 QKLKQQLMQMEREKAGLLATL---QDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQA 402
                | :|: |: ::: | |  :   :| : |  | | | :   : }  | : | |
Db        273 --LTKQIDQLSKDEVGKLTQIIQQKDLEIQALHARISSTSHTQDVVYLQQQLQA---YAM 327

Qy        403 SKERQTALDNEKDRD-SHEDGDYYE-VDINGPEILACKYHVAVAEAG--ELREQLKALRS 458
               :|:  |: ||| |: ||   :|:: :||           ||  ||   :|::: | | :
Db        328 EREKVFAVLNEKTRENSHLKTEYHKMMDI-----------VAAKEAALIKLQDENKKLST 376

Qy        459 THEA------RE-----AQHAEEKGRYEAEGQALTEK----VSLLEKA----------SR 493
               |:       ||      ::   ||   : |  ||::|    :::|: :          |
Db        377 RFESSGQDMFRETIQNLSRIIREK---DIEIDALSQKCQTLLAVLQTSSTGNEAGGVNSN 433

Qy        494 QDRELLA---RLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPN 550
               |  |||    :|::::||: :   :   ::   | |      |||  |   | : :::  :
Db        434 QFEELLQERDKLKQQVKKMEEWKQQVMTTVQNMQHESAQLQEELHQLQAQV-LVDSDNNS 492

Qy        551 RVMLDY------YREG--------------QGGAGRTSPGGRTSPEARGRR---SPILLP 587
               :: :||      | :                |   |:      |     |:    || |
Db        493 KLQVDYTGLIQSYEQNETKLKNFGQELAQVQHSIGQLC---NTKDLLLGKLDIISPQLSS 549

Qy        588 KGLLAPEAGRADGGTGDSSPSPGSS-LPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVD 646
               || |::      :     |  | |   | : |:        |  :::    |  ::
Db        550 ASLLTPQSAECLRASKSEVLSESSELLQQELEELRKSLQEKDATIRTLQENNHRLSDSIA 609

Qy        647 RTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL-STKREQITTLRT--VLKANK-Q 702
               |:|| |:     :      |:|: :::::|| | ||   | :|: :       | |: :
Db        610 ATSELERKEHEQTDSEIKQLKEKQDVLQKLLKEKDLLIKAKSDQLLSSNENFTNKVNENE 669

Qy        703 TAEVALANLKS-----------KYENEKAMVTETMMKLRNELKALKEDAATFSSL 746
                 |: |||             | ||||  : ||      | :|||:|    || :
Db        670 LLRQAVTNLKERILILEMDIGKLKGENEK--IVETYRGKETEYQALQETNMKFSMM 723


RESULT 14
US-10-146-473-82
; Sequence 82, Application US/10146473
; Publication No. US20030108888A1
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew
;  APPLICANT:  Gout, Ivan
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Gure, Ali
;  APPLICANT:  Chen, Yao-Tseng
;  APPLICANT:  Old, Lloyd
;  TITLE OF INVENTION: Breast Cancer Antigens
;  FILE REFERENCE: L00461/70130(JRV)
;  CURRENT APPLICATION NUMBER: US/10/146,473
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: US 60/291,150
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 82
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 82
;   LENGTH: 1388
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-146-473-82

  Query Match             7.2%;  Score 294;  DB 9;  Length 1388;
  Best Local Similarity   20.8%;  Pred. No. 1.8e-06;
  Matches  198;  Conservative  153;  Mismatches  343;  Indels  258;  Gaps   37;

Qy         12 RLVMEAQPEWLRAEVKRLSHEL---AETTREKIQ----AAEYGL---AVLEEKHQLKLQF 61
               :|  |::  :|  |  ||  ||    :| ||:|:     |:| :   :: ||  :|:|
Db        472 KLHKESRGGFLPEEQDRLLSELRNEIQTLREQIEHHPRVAKYAMENHSLREENRRLRL-L 530

Qy         62 EELEVDYEAIRSEMEQLKEAFGQAHTNHKK---------------VAADGESREESLIQ 105
               | ::   |     : :|::|| :     |                    : |: |  :  |:|
Db        531 EPVKRAQEMDAQTIAKLEKAFSEISGMEKSDKNQQGFSPKAQKEPCLFANTEKLKAQLLQ 590

Qy        106 ESA----SKEQYYV------RKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQN 155
                    ||::|        :: |||::||: |:    |   ||   |: | : :|::|
Db        591 IQTELNNSKQEYEEFKELTRKRQLELESELQSLQK--ANLNLENLLEATKACKRQEVSQL 648

Qy        156 VEIQRGRLR---DDIKEYKFREARL------LQDYSELEEENISLQKQVSVLRQ------ 200
               :|    |:      | |:      :  :  |  :| |:     :| :
Db        649 NKIHAETLKIITTPTKAYQLHSRPVPKLSPEMGSFGSLYTQNSSILDN-DILNEPVPPEM 707

Qy        201 NQVEFEGLKHEIKRLEE----------ETEYLNSQLEDAIRLKEISERQLE-----EALE 245
               |:  || :  |:: ::|          | |: | :|:  :   |    |::     | ::
Db        708 NEQAFEAISEELRTVQEQMSALQAKLDEEEHKNLKLQQHVDKLEHHSTQMQELFSSERID 767

Qy        246 TLKTERE---QKNSLRKELSHYMSINDSFYTSHLH----------VSLDGLKFSDDAAEP 292
                 | : |   | | | |:|    : || |  | :|           |  :|    : :
Db        768 WTKQQEELLSQLNVLEKQLQETQTKND-FLKSEVHDLRVVLHSADKELSSVKLEYSSFKT 826

Qy        293 NNDAE--ALVNGFEHGGLAKLPLDNKTSTPKK--EGLAPPSPSLVSDLLSELNISEIQKL 348
               | : |    |    |    :| |||     :|  |  |    |   | | |:   || :|
Db        827 NQEKEFNKLSERHMH---VQLQLDNLRLENEKLLESKACLQDSY--DNLQEIMKFEIDQL 881

Qy        349 KQQLMQMEREKAGLLATLQDTQKQLE-----HTRGSLSEQQEKVTRLTENLSALRRLQAS 403
               : |   ::|   | : | :  : ||      : : ||  :::|      | |  |  ::
Db        882 SRNLQNFKKENETLKSDLNNLMELLEAEKERNNKLSLQFEEDKENSSKEILKVLEAVRQE 941

Qy        404 KERQTA------------------------LDNEKDRDSHEDGDYYE------------ 426
               |:::||                      |:  :| |     |
Db        942 KQKETAKCEQQMAKVQKLEESLLATEKVISSLEKSRDSDKKVVADLMNQIQELRTSVCEK 1001

Qy        427 ---VDINGPEI--LACKYHVAVAEAGELREQLKA-------LRSTHEAREAQHAEEKGRY 474
                  :|    |:  : |||: |: :  | |  :|        |: |   |       |:
Db       1002 TETIDTLKQELKDINCKYNSALVDREESRVLIKKQEVDILDLKETLRLRILSEDIERDML 1061

Qy        475 EAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEEL 534
               :    ||:::: :||::   ||   ::|| |   :  | |  |:  ::|:    |:
Db       1062 CEDLAHATEQLNMLTEASKKHSGLLQSAQEELTKKEALIQELQHKLNQKKEEV----EQK 1117

Qy        535 ANLYH-------HVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLP 587
               | |:       ||     | |              : :| |  :|           |
Db       1118 KNEYNFKMRQLEHVMDSAAEDPQ------------SPKTPPHFQTH-----------LA 1153

Qy        588 KGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDR 647
               | |   |    ||    :|                                ::||
Db       1154 KLLETQEQEIEDGRASKTS----------------------------LEHL------ 1176

Qy        648 TTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEV- 706
               |:|                ::|:|    |||::|       ||:  :  :   ::|  |
Db       1177 VTKL--------------NEDREVKNAEILRMK-------EQLREMENLRLESQQLIEKN 1215

Qy        707 -----ALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQL 761
                    | ::| : ||      :       : :::||  |    :  |   : |     |
Db       1216 WLLQGQLDDIKRQKENSDQNHPDNQQLKNEQEESIKERLAKSKIVEEMLKMKAD-----L 1270

Qy        762 DEMQRQLAAAE-------DEKKTLNSLLRMAIQQKLALTQRLELLELDHEQT 806
```

```
Db       1271 EEVQSALYNKEMECLRMTDEVERTQTLESKAFQEKEQLRSKLEEMYEERERT 1322
```

RESULT 15
US-09-815-242-5229
; Sequence 5229, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5229
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5229

Query Match 7.1%; Score 292.5; DB 10; Length 1184;
Best Local Similarity 21.2%; Pred. No. 1.8e-06;
Matches 181; Conservative 162; Mismatches 328; Indels 183; Gaps 37;

```
Qy         39 EKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGES 98
Db        156 DRRQIIEESAGVLKYK---KRKAESL--------NKLDQTEDNL----TRVEDILYDLEG 200

Qy         99 REESLIQESASKEQYYVRKVLELQTELKQLRNVLT-----NTQSENERLASVAQELKEIN 153
Db        201 RVEPLKEEAAIAKEYKT-----LSHQMKHSDIVVTVHDIDQYTNDNRQLDQRLNDLQGQQ 255

Qy        154 QNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQK-------QVSVL---RQNQV 203
Db        256 ANKEADKQRLSQQIQQYKGKRHQLDNDVESLNYQLVKATEAFEKYTGQLNVLEERKKNQS 315

Qy        204 EFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSH 263
Db        316 ETNA------RYEEEQENLIELLENI--SNEISEAQ--DTYKSLKSKQKELNAVIRELEE 365

Qy        264 YMSINDSFYTSHL-HVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKK 322
Db        366 QLYVSDEAHDEKLEEIKNEYYTLMSEQSDVNNDIRFLKHTIEENEAKKSRLDSRLVE--- 422

Qy        323 EGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSE 382
Db        423 ----------VFEQLKDIQ-GQIKTTKKEYQQTNKELSAVDKEIKNIEKDLTDTKKAQNE 471

Qy        383 QQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDING-PEILACK--- 438
```

```
Db        472 YEEKL------YQAYRYTEKMKTRIDSLATQEEE--------YTYFFNGVKHILKAKNKE 517

Qy        439 ---YHVAVAEAGELREQLKALRSTHEAREAQH----AEEKGRYEAEGQALTE----KVSL 487
Db        518 LKGIHGAVAEIIDVPSKLTQAIETALGASLQHVIVDSEKDGRQAI--QFLKERNLGRATF 575

Qy        488 LEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNE 547
Db        576 LPLNVIQSRVVAT-------DIKSIAKEANGFISIA-SEAVKVAPEYQNI------IGNL 621

Qy        548 TPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSP 607
Db        622 LGNTIIVDHLKHANELA-----------RAIKYRTRIVTLEGDIVNPGGSMTGGGARKSK 670

Qy        608 SPGS------SLPSPLSDPRREPMNI---YNLIAI-----------------IRDQIKH 640
Db        671 SILSQKDELTTMRHQLEDYLRQTESFEQQFKELKIKSDQLSELYFEKSQKHNTLKEQVHH 730

Qy        641 LQAAVDR-TTELSRQRIASQEL------GPAVDKDKEALMEEILKLKSL-LSTKR--EQI 690
Db        731 FEMELDRLTTQETQIKNDHEEFEFEKNDGYTSDKSRQTLSEKETHLESIKASLKRLEDEI 790

Qy        691 TTLRTVLKANKQTAEVALANLKSKYENEKAMVTE-------TMMKLRN-------ELKAL 736
Db        791 ERYTKLSKEGKESVTKTQQTLHQK-QSDLAVVKERIKTQQQTIDRLNNQSQQTKHQLNDV 849

Qy        737 KEDAATFSSLRAMFATRCDEYITQ--LDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQ 794
Db        850 KEKIAFFNS---------DEVMGEQAFQNIKDQINGQQETRTRLSDELDKLKQQRIELNE 900

Qy        795 RLE----LLELDHE 804
Db        901 QIDAQEAILQVCHQ 914
```

Search completed: July 1, 2003, 08:58:07
Job time : 131.593 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 28.792 Seconds
(without alignments)
2751.276 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1330	32.5	782	2	A34219	Bic-D protein - fr
2	860	21.0	869	2	A88710	protein C43G2.2 [i
3	382	9.3	3225	2	I52300	giantin - human
4	382	9.3	3259	1	A56539	giantin - human
5	358.5	8.8	1938	1	A40997	myosin heavy chain
6	355	8.7	2139	2	T18296	myosin heavy chain
7	352.5	8.6	1790	2	S67593	transport protein
8	351.5	8.6	3187	2	JC5837	364K Golgi complex
9	348	8.5	1690	2	T13030	microtubule bindin
10	347.5	8.5	2007	1	B43402	myosin heavy chain
11	346.5	8.5	2116	2	A26655	myosin heavy chain
12	346	8.4	1976	2	A59252	myosin heavy chain
13	344.5	8.4	1972	1	A41604	myosin heavy chain
14	338.5	8.3	1979	1	S03166	myosin heavy chain
15	337.5	8.2	1927	2	A59236	embryonic muscle m
16	336.5	8.2	1938	2	JC5421	smooth muscle myos
17	335.5	8.2	1972	2	JC5420	smooth muscle myos
18	335	8.2	1269	2	F84730	probable myosin he
19	334	8.2	1175	2	D35815	myosin heavy chain
20	334	8.2	2057	2	S61477	myosin II heavy ch
21	333	8.1	1201	2	B35815	myosin heavy chain
22	331	8.1	1175	2	C35815	myosin heavy chain
23	331	8.1	1201	2	A35815	myosin heavy chain
24	331	8.1	2017	1	A36014	myosin heavy chain
25	330.5	8.1	1931	2	A59234	slow myosin heavy
26	329.5	8.0	1992	2	A47297	myosin heavy chain
27	329.5	8.0	2101	2	A42184	nuclear mitotic ap
28	329	8.0	1133	2	T22976	hypothetical prote
29	329	8.0	2253	2	T30336	nuclear/mitotic ap
30	328.5	8.0	1959	1	A33977	myosin heavy chain
31	328	8.0	1940	2	A59287	myosin heavy chain
32	328	8.0	1961	1	A61231	myosin heavy chain
33	327	8.0	2385	2	A32491	myosin heavy chain
34	327	8.0	2411	2	B32491	myosin heavy chain
35	326.5	8.0	1130	2	T34081	hypothetical prote
36	325	7.9	1955	2	T30934	myosin-like protei
37	325	7.9	2442	2	T08621	centrosome associa
38	325	7.9	2663	1	S28261	centromere protein
39	324	7.9	1909	2	A45592	liver stage antige
40	323.5	7.9	1957	2	T38077	hypothetical coile
41	321.5	7.9	1727	2	T50073	myosin-like coiled
42	321.5	7.9	1935	1	A37102	myosin beta heavy
43	320.5	7.8	1039	2	S18199	myosin heavy chain
44	320	7.8	1940	2	A29320	myosin heavy chain
45	320	7.8	1957	2	A59294	skeletal myosin -

ALIGNMENTS

RESULT 1
A34219
Bic-D protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 01-Dec-2000
C;Accession: A34219; A33636
R;Suter, B.; Romberg, L.M.; Steward, R.
Genes Dev. 3, 1957-1968, 1989
A;Title: Bicaudal-D, a Drosophila gene involved in developmental asymmetry: localized
A;Reference number: A34219; MUID:90152340; PMID:2576013
A;Accession: A34219
A;Molecule type: mRNA
A;Residues: 1-782 <SUT>
A;Cross-references: GB:X51652; NID:g7657; PIDN:CAA35964.1; PID:g7658
R;Wharton, R.P.; Struhl, G.
Cell 59, 881-892, 1989
A;Title: Structure of the Drosophila bicaudalD protein and its role in localizing the
A;Reference number: A33636; MUID:90075232; PMID:2590944
A;Accession: A33636
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-295,'S',297-317,'P',319-476,'R',478-782 <WHA>
A;Cross-references: GB:M31684; NID:g157005; PIDN:AAA28393.1; PID:g157006
C;Genetics:
A;Gene: FlyBase:BicD
A;Cross-references: FlyBase:FBgn0000183
C;Keywords: coiled coil

Query Match 32.5%; Score 1330; DB 2; Length 782;
Best Local Similarity 39.1%; Pred. No. 7.1e-47;
Matches 326; Conservative 148; Mismatches 247; Indels 112; Gaps 16;

```
Qy         22 LRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEA 81
              |: ||:||: || : :    |:|:|||::|||| |: : |||| |:  | |::  :||
Db         18 LQMEVERLTRELDQVSSASAQSAQYGLSLLEEKSALQQKCEELETLYDNTRHELDITQEA 77

Qy         82 FGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENER 141
                 :  |: |      |  :|:::|: |||::|     :: :|: ||||||: |   ::| :|
Db         78 LTKFQTSQKVTNKTGIEQEDALLNESAARETSLNLQIFDLENELKQLRHELERVRNERDR 137

Qy        142 LASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQN 201
                :     :      : |  | ||: ::|: |||| |:| :|||||||||||||||| || :
Db        138 MLQENSDFGRDKSDSEADRLRLKSELKDLKFRETRMLSEYSELEEENISLQKQVSSLRSS 197

Qy        202 QVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKEL 261
              |||||| ||||:|| || | || |:::   ||:|:|:|:|||||||: ||| | :|:|||
Db        198 QVEFEGAKHEIRRLTEEVELLNQQVDELANLKKIAEKQMEEALETLQGEREAKYALKKEL 257

Qy        262 SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPK 321
              :::    :: |:|  |:      :| |  |:| |      |:  | :|  |  |
Db        258 DGHLNRESMYHISNLAYSI--RSNMEDNASNNSDGEE-----ENLALKRLEADLSTELKS 310
```

```
Qy        322 KEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
Db        311 PDGTK-------CDLFSEIHLNELKKLEKQLESMESEKTHLTANLREAQTSLDKSQNELQ 363

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
Db        364 NFMSRLALLAAHVDALVQL------------KKQIDVKEQG------------------- 392

Qy        442 AVAEAG----ELREQLKALRS------THEAREAQHAEEKGRYEAEGQALTEKVSLL-EK 490
Db        393 --KEGGQKKDELEQQLRALISQYANWFTLSAKEIDGLKTDIAELQKGLNYTDATTTLRNE 450

Qy        491 ASRQDRELLARLEKELKKVSDV---------AGETQGSLSVAQDELVTFSEELANLYHHV 541
Db        451 VTNLKNKLLATEQKSLDLQSDVQTLTHISQNAGQSLGS---ARSTLVALSDDLAQLYHLV 507

Qy        542 CMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGG 601
Db        508 CTVNGETPTRVLLDHKTDDMS----FENDSLTAIQSQFKSDVFIAKPQIVEDLQGLADS- 562

Qy        602 TGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQEL 661
Db        563 ----------------------VEIKKYVDTVSDQIKYLKTAVEHTIDMNKHKIRS-EG 598

Qy        662 GPAVDK----DKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYEN 717
Db        599 GDALEKVNTEEMEELQEQIVKLKSLLSVKREQIGTLRNVLKSNKQTAEVALTNLKSKYEN 658

Qy        718 EKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKT 777
Db        659 EKIIVSDTMSKLRNELRLLKEDAATFSSLRAMFAARCEEYVTQVDDLNRQLEAAEEEKKT 718

Qy        778 LNSLLRMAIQQKLALTQRLELLELDHEQ--------TRRGRA-KAAPKTKPAT 821
Db        719 LNQLLRLAVQQKLALTQRLEEMEMDREMRHVRRPMPAQRGTSGKSSFSTRPSS 771


RESULT 2
A88710
protein C43G2.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: A88710
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-869 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAB09108.1; PID:g1572757; GSPDB:GN00022; CESP:C43G2.
C;Genetics:
A;Gene: C43G2.2
A;Map position: 4

  Query Match          21.0%;  Score 860;  DB 2;  Length 869;
  Best Local Similarity   30.3%;  Pred. No. 8.3e-28;
  Matches  255;  Conservative  144;  Mismatches  260;  Indels  182;  Gaps   20;

Qy         16 EAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEM 75
Db          3 ESELEKLRQDIAILTEKYEQAKEDIHKAANAGLELLRQKEDLEKRLAEMQAELDLARTEI 62

Qy         76 EQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNT 135
Db         63 DKTNQTLAEYRSQHQRSTRSELENEESLLEESSAKEEEYLQRIAKLEADLKKKEQELAEK 122

Qy        136 QSENERLASVAQELKEINQN--VEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQK 193
Db        123 KEELESIEK--KHSKEIDSGAALEDERRKLRAELKETKEREQRLISEYSELEEENIGLQK 180

Qy        194 QVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQ 253
Db        181 TVANLRGSQVEYESLRIDNNRLEETIEIMKMAAEEDEILRVIADKQLEEALLTAQQERDQ 240

Qy        254 KNSLRKEL------SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGG 307
Db        241 RLAMKRELEQTRNAEHISSLNDMLF---------------------------------G 266

Qy        308 LAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQ 367
Db        267 LERLGEDGE--------LPPPQPG-ASDLFSELQGSSDVKVR----ELEAAKEGLQEELK 313

Qy        368 DTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEV 427
Db        314 SREKIF----------IEFVTGLADTLNIHRP----------TNELD------------ 340

Qy        428 DINGPEILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALTEKVSL 487
Db        341 ----------YMHARQQKDVVLEKIQNI----------ARDTDRHDKEGE---EKRSG 375

Qy        488 LEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH------- 540
Db        376 ILKA------------DLRTLVLVAGEKSAQLAAAQDAMIQVSDQLYQFYHQMTQNQGV 422

Qy        541 ---------------VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPIL 585
Db        423 QTEKSVQEIVKKLRLLARANAEDVPRVSL---------ADEGVESGTETDVNASRSIPLN 473

Qy        586 LPKGLLAP------EAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIK 639
Db        474 SDRLVIAPSFAKEIEKKLASVKIGDVLSE---------TDLRQRILTEGNAISETTESLK 524

Qy        640 HLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKA 699
Db        525 KMIQVVKRTSEQ-----AFNQAVMASGAENEIEMQN-MKLRSLLSTKRDQISTLRTVLKS 578

Qy        700 NKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYIT 759
Db        579 NKLTAESALTSMREKYESEKKMMMEINDKMRRELKQLKEDAATFASHRAMFTARGEELKS 638

Qy        760 QLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQT--RRGRAKAAPKT 817
Db        639 KVEELSNELRANEEEKKTLNQLLRLAIQQKLTLTQRLEEVEVDRDRQVFKRSSTRAPTRE 698

Qy        818 K 818
Db        699 K 699


RESULT 3
I52300
giantin - human
N;Alternate names: gcp372
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C;Accession: I52300
R;Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A;Title: Molecular cloning and sequence analysis of a human 372-kDA protein localized
A;Reference number: I52300; MUID:95100974; PMID:7802676
A;Accession: I52300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3225 <RES>
A;Cross-references: GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808869
C;Superfamily: giantin

  Query Match          9.3%;  Score 382;  DB 2;  Length 3225;
  Best Local Similarity   21.8%;  Pred. No. 8.9e-08;
  Matches  205;  Conservative  179;  Mismatches  361;  Indels  196;  Gaps   34;

Qy          5 SEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLE------------ 52
```

```
Db       1971 SHAKELQELLKEKQ-----QEVKQLQKDCIR-YQEKISALERTVKALEFVQTESQKDLEI 2024

Qy         53 --------EKHQLKLQFE--------------------------ELEVDYEAIRSEMEQ 77
                    :|: | | |                             ||: : |:::|:|:|
Db       2025 TKENLAQAVEHRKKAQAELASFKVLLDDTQSEAARVLADNLKLKKELQSNKESVKSQMKQ 2084

Qy         78 ----LKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLT 133
                   |:    ||   | |   : : : ::| :|    |:       : |  : |::: :
Db       2085 KDEDLERRLEQAEEKHLKEKKNMQEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQE 2144

Qy        134 NTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSE---LEEENIS 190
               |  |   :||:  : :  :  :    | |: |: |::: : :  :|   |   |:|:| |
Db       2145 NLDSTVTQLAAFTKSMSSLQDD----RDRVIDEAKKWERKFSDAIQSKEEEIRLKEDNCS 2200

Qy        191 LQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKE--------ISERQLEE 242
               : |    |||  :  | ||  | |||  : :   |: :  ::|::         ::  | :
Db       2201 VLK--DQLRQMSIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLSQ 2258

Qy        243 ALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNG 302
                 ||     :| | |  |   |: |    : |  ||:  |       |   : | ::
Db       2259 LEETRHLYHSSQNELAKLESELKSLKDQL--TDLSNSLEKCK------EQKGNLEGIIRQ 2310

Qy        303 FEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGL 362
               |    | :  ::| |  : |     |  | | |  |:|: | ||:   |   |
Db       2311 QE----ADIQ-NSKFSYEQLETDLQASRELTSRLHEEINMKE-QKIISLLSGKEEAIQVA 2364

Qy        363 LATL-QDTQKQLEHTRGSLSEQQE-------------KVTRLTENLSALRRLQASKERQ 407
               :| | |   |:::    ||:::|             |  :| | |  :::    :: |
Db       2365 IAELRQQHDKEIKELENLLSQEEEENIVLEEENKKAVDKTNQLMETLKTIKKENIQQKAQ 2424

Qy        408 --------TALDNEKDRDSHEDGDYYEVDINGPEILACKYHV---AVAEAGELREQLKAL 456
                     ::| |::||     ||| :::     |:  |  :   | ||  :|:|::: |
Db       2425 LDSFVKSMSSLQNDRDRIV---GDYQQLEERHLSIILEKDQLIQEAAAENNKLKEEIRGL 2481

Qy        457 RSTHEAREAQHAE---EKGRYEAE-GQALTEKVS----LLEKASRQDREL---LARLEKE 505
               ||  :   :::|:   |  :|  :  | :| | |    |||   :|::||    |:|||::
Db       2482 RSHMDDLNSENAKLDAELIQYREDLNQVITIKDSQQKQLLEVQLQQNKELENKYAKLEEK 2541

Qy        506 LKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565
               ||:  :   : : | :  |:|    |:|: :|  :      :             : :|  |
Db       2542 LKESEEANEDLRRSFNALQEEKQDLSKEIESLKVSISQLTRQVT-------ALQEEGTLG 2594

Qy        566 RTSPGGRTSPEARGRRSPIL---------LPKGLLA--PEAGRADGGTGDSSPSPGSSLP 614
                  :   |   | | :          | : |:    || :  |   |
Db       2595 LYHAQLKVKEEEVHRLSALFSSSQKRIAELEEELVCVQKEAAKKVGEIEDKL-------- 2646

Qy        615 SPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALME 674
                 ::|  ::::   |:|::    : | :|  ||:|  :  ::    | |: : |
Db       2647 ------KKELKHLHHDAGIMRNE---TETAEERVAELARDLVEMEQKLLMVTKENKGLTA 2697

Qy        675 EILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYE---NEKAMVTETMMKLRN 731
               :|      :|:           |: ::   |    |  || ||:    | | : |  :  |
Db       2698 QIQSFGRSMSS-----------LQNSRDHANEELDELKRKYDASLKELAQLKEQGLLNRE 2746

Qy        732 ELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLA 791
               | | | : :|        :  :: |::: :|| : :::    |:| |  :  |  :
Db       2747 RDALLSETAFSMNSTE-------ENSLSHLEKLNQQLLSKDEQLLHLSSQLEDSYNQVQS 2799

Qy        792 LTQRLELL--ELDH-----EQTRRG-----RAKAAPKTKPA 820
                :: :  |  | ||     |: |:      |: | | | ||
Db       2800 FSKAMASLQNERDHLWNELEKFRKSEEGKQRSAAQPSTSPA 2840


RESULT 4
A56539
giantin - human
N;Alternate names: macrogolgin
C;Species: Homo sapiens (man)
C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C;Accession: A56539; S37536
R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994

A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protei
A;Reference number: A56539; MUID:94187728; PMID:7511208
A;Accession: A56539
A;Molecule type: mRNA
A;Residues: 1-3259 <SEE>
A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C;Genetics:
A;Gene: GDB:GOLGB1; GCP; GCP371
A;Cross-references: GDB:454958
A;Map position: 3q13.31-3q13.31
C;Superfamily: giantin
C;Keywords: coiled coil; Golgi apparatus; transmembrane protein
F;3238-3254/Domain: transmembrane #status predicted <TMN>

  Query Match             9.3%;  Score 382;  DB 1;  Length 3259;
  Best Local Similarity.  21.8%;  Pred. No. 9e-08;
  Matches  205;  Conservative  179;  Mismatches  361;  Indels  196;  Gaps   34;

Qy          5 SEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLE----------- 52
                | :|   |: | |       |||:|  :      :||| | |  :  ||
Db       2005 SHAKELQELLKEKQ-----QEVKQLQKDCIR-YQEKISALERTVKALEFVQTESQKDLEI 2058

Qy         53 --------EKHQLKLQFE--------------------------ELEVDYEAIRSEMEQ 77
                     :|: | | |                             ||: : |:::|:|:|
Db       2059 TKENLAQAVEHRKKAQAELASFKVLLDDTQSEAARVLADNLKLKKELQSNKESVKSQMKQ 2118

Qy         78 ----LKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLT 133
                   |:    ||   | |   : : : ::| :|    |:       : |  : |::: :
Db       2119 KDEDLERRLEQAEEKHLKEKKNMQEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQE 2178

Qy        134 NTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSE---LEEENIS 190
               |  |   :||:  : :  :  :    | |: |: |::: : :  :|   |   |:|:| |
Db       2179 NLDSTVTQLAAFTKSMSSLQDD----RDRVIDEAKKWERKFSDAIQSKEEEIRLKEDNCS 2234

Qy        191 LQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKE--------ISERQLEE 242
               : |    |||  :  | ||  | |||  : :   |: :  ::|::         ::  | :
Db       2235 VLK--DQLRQMSIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLSQ 2292

Qy        243 ALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNG 302
                 ||     :| | |  |   |: |    : |  ||:  |       |   : | ::
Db       2293 LEETRHLYHSSQNELAKLESELKSLKDQL--TDLSNSLEKCK------EQKGNLEGIIRQ 2344

Qy        303 FEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGL 362
               |    | :  ::| |  : |     |  | | |  |:|: | ||:   |   |
Db       2345 QE----ADIQ-NSKFSYEQLETDLQASRELTSRLHEEINMKE-QKIISLLSGKEEAIQVA 2398

Qy        363 LATL-QDTQKQLEHTRGSLSEQQE-------------KVTRLTENLSALRRLQASKERQ 407
               :| | |   |:::    ||:::|             |  :| | |  :::    :: |
Db       2399 IAELRQQHDKEIKELENLLSQEEEENIVLEEENKKAVDKTNQLMETLKTIKKENIQQKAQ 2458

Qy        408 --------TALDNEKDRDSHEDGDYYEVDINGPEILACKYHV---AVAEAGELREQLKAL 456
                     ::| |::||     ||| :::     |:  |  :   | ||  :|:|::: |
Db       2459 LDSFVKSMSSLQNDRDRIV---GDYQQLEERHLSIILEKDQLIQEAAAENNKLKEEIRGL 2515

Qy        457 RSTHEAREAQHAE---EKGRYEAE-GQALTEKVS----LLEKASRQDREL---LARLEKE 505
               ||  :   :::|:   |  :|  :  | :| | |    |||   :|::||    |:|||::
Db       2516 RSHMDDLNSENAKLDAELIQYREDLNQVITIKDSQQKQLLEVQLQQNKELENKYAKLEEK 2575

Qy        506 LKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565
               ||:  :   : : | :  |:|    |:|: :|  :      :             : :|  |
Db       2576 LKESEEANEDLRRSFNALQEEKQDLSKEIESLKVSISQLTRQVT-------ALQEEGTLG 2628

Qy        566 RTSPGGRTSPEARGRRSPIL---------LPKGLLA--PEAGRADGGTGDSSPSPGSSLP 614
                  :   |   | | :          | : |:    || :  |   |
Db       2629 LYHAQLKVKEEEVHRLSALFSSSQKRIAELEEELVCVQKEAAKKVGEIEDKL-------- 2680

Qy        615 SPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALME 674
                 ::|  ::::   |:|::    : | :|  ||:|  :  ::    | |: : |
Db       2681 ------KKELKHLHHDAGIMRNE---TETAEERVAELARDLVEMEQKLLMVTKENKGLTA 2731

Qy        675 EILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYE---NEKAMVTETMMKLRN 731
               :|      :|:           |: ::   |    |  || ||:    | | : |  :  |
Db       2732 QIQSFGRSMSS-----------LQNSRDHANEELDELKRKYDASLKELAQLKEQGLLNRE 2780
```

```
Qy        732 ELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLA 791
Db       2781 RDALLSETAFSMNSTE-------ENSLSHLEKLNQQLLSKDEQLLHLSSQLEDSYNQVQS 2833

Qy        792 LTQRLELL--ELDH-----EQTRRG-----RAKAAPKTKPA 820
Db       2834 FSKAMASLQNERDHLWNELEKFRKSEEGKQRSAAQPSTSPA 2874


RESULT 5
A40997
myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Aequipecten irradians
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: A40997; S13557
R;Nyitray, L.; Goodwin, E.B.; Szent-Gyoergyi, A.G.
J. Biol. Chem. 266, 18469-18476, 1991
A;Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq
A;Reference number: A40997; MUID:92011595; PMID:1917970
A;Accession: A40997
A;Molecule type: mRNA
A;Residues: 1-1938 <NYI>
A;Cross-references: GB:X55714; NID:g5611; PIDN:CAA39247.1; PID:g5612
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b
F;86-763/Domain: myosin motor domain homology <MMOT>
F;176-183/Region: nucleotide-binding motif A (P-loop)
F;547-586/Region: actin binding #status predicted
F;653-675/Region: actin binding #status predicted
F;836-1938/Domain: coiled coil #status predicted <COI>
F;836-1276/Region: S2
F;1277-1938/Region: light meromyosin
F;182/Binding site: ATP (Lys) #status predicted
F;693,703/Active site: Cys #status predicted

  Query Match             8.8%;  Score 358.5;  DB 1;  Length 1938;
  Best Local Similarity   21.7%;  Pred. No. 4.3e-07;
  Matches  197;  Conservative  155;  Mismatches  352;  Indels  203;  Gaps   32;

Qy         13 LVMEAQPEWLRAEVK---RLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELE---- 65
Db        837 LSIARQEEEMKEQLKQMDKMKEDLAKTERIKKELEEQNVTLLEQKNDLFLQLQTLEDSMG 896

Qy         66 ---------------------------------VDYEAIRSEME----QLKEAFGQAHT 87
Db        897 DQEERVEKLIMQKADFESQIKELEERLLDEEDAAADLEGIKKKMEADNANLKKDIGDLEN 956

Qy         88 NHKKVAADGESREE--SLIQESASKEQYYVRKVLELQTELKQLRNVLTNT----QSE--- 138
Db        957 TLQKAEQDKAHKDNQISTLQGEISQQDEHIGK---LNKEKKALEEANKKTSDSLQAEEDK 1013

Qy        139 ----NERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQ 194
Db       1014 CNHLNKLKAKLEQALDELEDNLEREK-KVRGDVEKAK----------RKVEQDLKSTQEN 1062

Qy        195 VSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRL-----KEISERQ--LEEALETL 247
Db       1063 VEDLERVKRE---LEENVRRKEAEISSLNSKLEDEQNLVSQLQRKIKELQARIEELEEEL 1119

Qy        248 KTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGG 307
Db       1120 EAERNARAKVEKQ-----------RAELNRELEELGERLDEAGGATSAQIELNKKREAE 1167

Qy        308 LAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSEL--NISEIQKLKQQLMQMEREKAGLLAT 365
Db       1168 LLKIRRDLEEASLQHEAQISALRKKHQDAANEMADQVDQLQKVKSKL---EKDKKDLKRE 1224

Qy        366 LQDTQKQLEHTRGSLSEQQEKVTRLTENLSALR-RLQASKERQTALDNEKDRDSHEDGDY 424
Db       1225 MDDLESQMTHNMKNKGCSEKVMKQFESQMSDLNARLEDSQRSINELQSQKSRLQAENSDL 1284

Qy        425 YEVDINGPEILACKYHVAV--AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALT 482
Db       1285 TR------QLEDAEHRVSVLSKEKSQLSSQLE------DARRSLEEETRARSKLQNEVRN 1332

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDEL----VTFSEELANLY 538
Db       1333 MHADM--DAIRE------QLEEEQESKSDV----QRQLSKANNEIQQWRSKFESEGANRT 1380

Qy        539 HHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPE---- 594
Db       1381 EEL-----EDQKRKLL----------GKLSEAEQTTEAANAKCSALEKAKSRLQQELEDM 1425

Qy        595 ----------------AGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQI 638
Db       1426 SIEVDRANASVNQMEKKQRAFDKTTAEWQAKVNSLQSELENSQKESRGYSAELYRIKASI 1485

Qy        639 KHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLS---TKREQITTLRT 695
Db       1486 EEYQDSIG----------------ALRRENKNLADEIHDLTDQLSEGGRSTHELDKARR 1528

Qy        696 VLKANKQTAEVALANLKSKYENEKAMVTETMMKL---RNEL-KALKEDAATFSSLRAMFA 751
Db       1529 RLEMEKEELQAALEEAEGALEQEEAKVMRAQLEIATVRNEIDKRIQEKEEEFDNTRR--- 1585

Qy        752 TRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRA 811
Db       1586 ----NHQRALESMQASLEAEAKGKAD-------AMRIKKKLEQDINELEVALDASNRGKA 1634

Qy        812 KAAPKTK 818
Db       1635 EMEKTVK 1641


RESULT 6
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

  Query Match             8.7%;  Score 355;  DB 2;  Length 2139;
  Best Local Similarity   22.0%;  Pred. No. 6.7e-07;
  Matches  192;  Conservative  145;  Mismatches  315;  Indels  222;  Gaps   34;

Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQ----- 60
Db       1221 DETEVEKEDAEKKKKEIEKEMKALQEE-----KENVESSKNS----TEKDKKKLEDNLKD 1271

Qy         61 ----FEELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVR 116
Db       1272 TQKKLDDMTADNEKLKAKAKDLEAQLNEVQDNHEKAVADAELLNKKKAQ--SDKELNSLK 1329

Qy        117 KVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQ---NVEIQRGRLRDDIKEYKFR 173
Db       1330 AELEALTKAKSVVE-SKNKDSENEK-AALSEEIDQANEKLKNIQADLRKATADLQEANEK 1387

Qy        174 EARLLQDYSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLK 233
Db       1388 KAEVEAQRDKLVADNKKMTKTL-------------EEIKARDEENTYKVENYEKVLKRK 1433

Qy        234 EISERQLEEALETLKTEREQKNSLRKE-------------LSHYMSINDSFYTSHLHVS 279
```

```
Db       1434 ---EADLEEANENLDIEKKDRMNKEKQVKKLEGELKETKDKLNAAIAEKDSIFTAKKQSD 1490

Qy        280 LDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGL---APPSPSLVSDL 336

Db       1491 AD-LEELNKTVEEHDEVVAKLN----TQITKLTRDNQSAEEELNELRSKADKDKKKISEL 1545

Qy        337 LSELN-------------ISEIQKLKQQLMQMER--EKAGLLAT-LQDTQKQLEHTRGSL 380

Db       1546 EEQVNELESRPVGTGNADENEIKIRDAQIADLNKALEMKGVQNNQLQATNKELKAKDNDL 1605

Qy        381 SEQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYH 440

Db       1606 TSKIE----ITEN--EMKKLENAKKR---LEQDKD----------EAD------------ 1634

Qy        441 VAVAE----AGELREQLKALRSTHEAREAQ-HAEEKGRYEAEGQALTEKVSLLEKASRQD 495

Db       1635 KAVSEQTIKRKGLEEEVKKLTTEIQALKFQINAPSSVAQEEEKQRLESDIAELKEQLEQE 1694

Qy        496 RELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLD 555

Db       1695 RTTAANAEAERKKI-------QAEL----DEVKFNLEDVTNQREKLVAKNSE--NDAEID 1741

Qy        556 YYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPS 615

Db       1742 SLKE------------------------------------------------------- 1745

Qy        616 PLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKD---KEAL 672

Db       1746 -------EKKALEDEIEKITDDNNKLSEEID---SLDRKYNA---LLDSKDSDVSMKEKF 1792

Qy        673 MEEILKLKSLLST-KREQITTLRTVLKANKQTAEV-----ALANLKSKYENEKAMVTETM 726

Db       1793 QDELKVTKDALETEKKNHAETMRLKGRLEKEAAEVQVRLEALQKNLDLAQQEKAKATKDY 1852

Qy        727 MKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786

Db       1853 RAADGELKSLMNE-------------LDDVKDQLDKAQDDLADKEDELATLDQKYKTLV 1898

Qy        787 QQKLALTQRLELLELDHEQTRRGRAKAAPKTKPA 820

Db       1899 KQKSVFDSRIQEMQEQLDLEKAGRAK-AQKQKQA 1931


RESULT 7
S67593
transport protein USO1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D2552; protein YDL058w
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: S67593; A38455; S30782
R;Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67587
A;Accession: S67593
A;Molecule type: DNA
A;Residues: 1-1790 <BLO>
A;Cross-references: EMBL:Z74106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w
A;Experimental source: strain S288C
R;Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A;Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transp
A;Reference number: A38455; MUID:91185402; PMID:2010462
A;Accession: A38455
A;Molecule type: DNA
A;Residues: 1-389,'TA',392-724,'S',726-1790 <NAK>
A;Cross-references: GB:X54378; NID:g4777; PIDN:CAA38253.1; PID:g4778
A;Note: the authors translated the codon ACT for residue 768 as Ile
R;Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A;Description: An integrin analogue in Saccharomyces cerevisiae.
A;Reference number: S30782
A;Accession: S30782
A;Molecule type: DNA
A;Residues: 71-846,'E',848-923,'K',925-1252,'I',1254-1318,'V',1320-1460,'S',1462-1580
A;Cross-references: EMBL:L03188
C;Genetics:
A;Gene: SGD:USO1; INT1
A;Cross-references: SGD:S0002216; MIPS:YDL058w
A;Map position: 4L
C;Keywords: coiled coil; transmembrane protein
F;326-342/Domain: transmembrane #status predicted <TM1>
F;394-410/Domain: transmembrane #status predicted <TM2>
F;617-633/Domain: transmembrane #status predicted <TM3>

  Query Match             8.6%;  Score 352.5;  DB 2;  Length 1790;
  Best Local Similarity   20.5%;  Pred. No. 6.8e-07;
  Matches  193;  Conservative  176;  Mismatches  307;  Indels  266;  Gaps   34;

Qy         25 EVKRLSHELAETTREKIQAAEYGL-AVLEEKHQLKLQFEELEVDYEAIRSEME------- 76

Db        829 EYKSTIHK----QEDSIKTLEKGLETILSQKKKAEDGINKMGKDLFALSREMQAVEENCK 884

Qy         77 QLKEAFGQAHTNHKK--------------VAADGESREESLIQ-ESASKEQYYVRK--- 117

Db        885 NLQKEKDKSNVNHQKETKSLKEDIAAKITEIKAINENLEEMKIQCNNLSKEKEHISKELV 944

Qy        118 ------------VLELQTELKQLRNVLTNTQSENERLASVAQELK-------------- 150

Db        945 EYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAENESLIKAVEESKNESSIQLSNLQNKID 1004

Qy        151 ---EINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENIS--------LQKQVSVLR 199

Db       1005 SMSQEKENFQIERGSIEKNIEQLK----KTISDLEQTKEEIISKSDSSKDEYESQISLLK 1060

Qy        200 Q-------------NQVE--------------FEGLKHEIK-RLEEETEYLNSQLEDAI 230

Db       1061 EKLETATTANDENVNKISELTKTREELEAELAAYKNLKNELETKLETSEKALKEVKENEE 1120

Qy        231 RLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAA 290

Db       1121 HLKEEKIQLEKEATET----KQQLNSLRANL-------ESLEKEHEDLAAQLKKYEEQIA 1169

Qy        291 EPN---NDAEALVNGFEHGGLAKLPLDNKTSTPKK-EGLAPPSPSL---------VSDLL 337

Db       1170 NKERQYNEEISQLN-----------DEITSTQQENESIKKKNDELEGEVKAMKSTSEEQ 1217

Qy        338 SELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSAL 397

Db       1218 SNLKKSEIDALNLQIKELKKKNETNEASLLESIKSVESETVKIKELQDECNFKEKEVSEL 1277

Qy        398 R-RLQASKERQTA-LDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKA 455

Db       1278 EDKLKASEDKNSKYLELQKESEKIKE----ELD--------------AKTTELKIQLEK 1318

Qy        456 LRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDR----ELLARLEKELKKVSD 511

Db       1319 ITNLSKAKEKSESE--------------LSRLKKTSSEERKNAEEQLEKLKNEI-QIKN 1362

Qy        512 VAGE------TQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565

Db       1363 QAFEKERKLLNEGSSTITQE----YSEKINTLEDELIRLQNENELK-------------- 1404

Qy        566 RTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPM 625

Db       1405 -----AKEIDNTRSELEKVSLSNDELLEEKQNTIKSLQDEI----------LSYKDKITR 1449

Qy        626 NIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLST 685

Db       1450 NDEKLLSIERDNKRDLES-------LKEQLRAAQESKAKVEEGLKKLEEESSKEKAELEK 1502

Qy        686 KREQITTLRTVLKAN--------------------KQTAEVALANLK-------SKYEN 717

Db       1503 SKEMMKKLESTIESNETELKSSMETIRKSDEKLEQSKKSAEEDIKNLQHEKSDLISRINE 1562

Qy        718 EKAMVTETMMKLR------NELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAA 771
```

```
Db       1563 SEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINAEENTVLKSKLEDIERELKDK 1622

Qy        772 EDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKA 813
Db       1623 QAEIKSNQ-------EEKELLTSRLKELEQELDSTQQKAQKS 1657


RESULT 8
JC5837
364K Golgi complex-associated protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C;Accession: JC5837
R;Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A;Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A;Reference number: JC5837; MUID:98093490; PMID:9431462
A;Accession: JC5837
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3187 <TOK>
A;Cross-references: DDBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:g516826
C;Comment: This protein plays a role in the formation and maintenance of the characteris
C;Superfamily: giantin
F;49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F;3165-3187/Domain: membrane anchor #status predicted <MAD>

  Query Match             8.6%;  Score 351.5;  DB 2;  Length 3187;
  Best Local Similarity   22.2%;  Pred. No. 1.5e-06;
  Matches  199;  Conservative  178;  Mismatches  350;  Indels  171;  Gaps   36;

Qy         22 LRAEVKRLSHELAETTREK-IQAAEYGLAVLEEKHQ------------------------ 56
Db       1974 LEKTVKAL--EFVHTESQKDLDATKGNLAQAVEHHKKAQAELSSFKILLDDTQSEAARVL 2031

Qy         57 ---LKLQFEELEVDYEAIRSEMEQ----LKEAFGQAHTNHKKVAADGESREESLIQESAS 109
Db       2032 ADNLKLK-KELQSNKESIKSQIKQKDEDLLRRLEQAEEKHRKEKKNMQEKLDALHREKAH 2090

Qy        110 KEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKE 169
Db       2091 VEDTLAEIQVSLTRKDKDMKEL---QQSLDSTLAQLAAFTKSMS-SLQDDRDRVIDEAKK 2146

Qy        170 YKFREARLLQDYSE---LEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQL 226
Db       2147 WEQRFGDAIQTKEEEVRLKEENCTALK--DQLRQMTIHMEELKITVSRLEHDKEIWESKA 2204

Qy        227 EDAIRLKEISERQLEEALETLKTEREQ--------KNSLRKELSHYMSINDSFYTSHLHV 278
Db       2205 QTELQHQQKAYDKLQEENKELMSQLEEAGQLYHDSKNELTKLESELKSLKDQ--STDLKN 2262

Qy        279 SLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDN-KTSTPKKEGLAPPSPSLVSDLL 337
Db       2263 SLEKCR------EHENNLEGIIKQQE------ADIQNCKFNCEQLETDLTASRELTTRLH 2310

Qy        338 SELNISEIQKLKQQLMQMEREKAGLLATL-QDTQKQLEHTRGSLSEQQ------------ 384
Db       2311 DEINVKE-QKIISLLSGKEEAIQVAIAELHQQHSKEIKELENLLSQEEEENLTLEEENKR 2369

Qy        385 --EKVTRLTENLSALRRLQASKERQTALDN--------EKDRDSHEDGDYYEVDINGPEI 434
Db       2370 AVEKTNQLTEALETIKK--ESLEQKAQLDSFVKSMSSLQDDRD-RIVSDYRQLEERHLSV 2426

Qy        435 LACKYHV---AVAEAGELREQLKALRSTHEAREAQHAE---EKGRYEAE-GQALTEKVS- 486
Db       2427 ILEKDELIQDAAAENNKLKEEIRGLRGHMDDLNSENAKLDAELIQYRRDLNEVITIKDSQ 2486

Qy        487 ---LLEKASRQDREL---LARLEKELKKVSDVAGETQGSLSVAQD-------ELVTFSEE 533
Db       2487 QRQLLEAQLQQNKELRNECVKLEGRLKGSEAEKQSLQMSLDALQEENQGLSKEIKSFKEQ 2546

Qy        534 LANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAP 593
```

```
Db       2547 LTALHEEGALAVYHAQLRVREEEVQK-----------LTAALSSSQKRTVDLQEELVCV 2594

Qy        594 EAGRADGGTGDSSPSPGSSLPSPLSDP-RREPMNIYNLIAIIRDQIKHLQAAVDRTTELS 652
Db       2595 Q------------KEASKKVSEIEDKLKRELKHLHHNAGIMRNE---TETAEERVAELA 2638

Qy        653 RQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLK 712
Db       2639 RDLVEMEQKLLTVTKENKDLTAQIQAFGKSMSS----------LQDSRDHATEELSDLK 2687

Qy        713 SKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAMFA---TRCDEYITQLDEMQRQLA 769
Db       2688 KKY--------DASLKELAQLKG-RQDLGRESDVLSQAAFPLTTSENISSRLEKLNQQLI 2738

Qy        770 AAEDEKKTLNSLLRMAIQQKLALTQRLELL--ELDH--EQTRRGRAKAAPKTKPATPS 823
Db       2739 SKDEQLLHLSSELESSHNQVQSFTKAMTSLQNERDHLWNELEKFRKSEEGKQRSAAPS 2796


RESULT 9
T13030
microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13030
R;Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A;Title: A class VI unconventional myosin is associated with a homologue of a microtu
A;Reference number: Z17588; MUID:98139549; PMID:9472041
A;Accession: T13030
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1690 <LAN>
A;Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1
A;Experimental source: strain Oregon R
C;Genetics:
A;Cross-references: FlyBase:FBgn0020503
C;Keywords: cytoskeleton

  Query Match             8.5%;  Score 348;  DB 2;  Length 1690;
  Best Local Similarity   20.1%;  Pred. No. 9.7e-07;
  Matches  181;  Conservative  184;  Mismatches  329;  Indels  206;  Gaps   34;

Qy          5 SEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAV------LEEK---- 54
Db        741 NELEDFQK--KQSESEVHLQEIK------AQNTQKDLELVESGESLKKLQQQLEEKTLGH 792

Qy         55 HQLKLQFEELEVDYEAI----RSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASK 110
Db        793 EKLQAALEELKKEKETIIKEKEQELQQLQSKSAESESALKVV----QVQLEQLQQQAAAS 848

Qy        111 EQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEY 170
Db        849 GEEGSKTVAKLHDEISQLKSQAEETQSE---LKSTESNLEAKSKQLEAANGSLEEEAKKS 905

Qy        171 KFREARLLQDYSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDA- 229
Db        906 GQLQEQITKLKSEVEETQAALSSYHTDVESKTKQLEAANAALEKVNK--EYAESRAEASD 963

Qy        230 --IRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSD 287
Db        964 LQDKVKEITDTLHAE----LQAERSSSSALHTKLSKF---SDEIATGHKELTSKADAWSQ 1016

Qy        288 DAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQK 347
Db       1017 EMLQKEKELQELRQQLQ---------DSQDSQTKLKAEGERKEKSFEESIKNLQ-EEVTK 1066

Qy        348 LKQQLMQMEREKAGLLATLQDTQKQLEHTRGSL-------SEQQEKVTRLTENLSALRRL 400
Db       1067 AKTENLEL---STGTQTTIKDLQERLEITNAELQHKEKMASEDAQKIADLKTLVEAIQVA 1123

Qy        401 QASKERQTA--------LDNEKDRDSHEDGDY-YEVDINGPEILACKYHVAVAEAGELRE 451
```

```
Db       1124 NANISATNAELSTVLEVLQAEKSETNHIFELFEMEADMNSERL--------IEKVTGIKE 1175

Qy        452 QLKALRSTH--------------EAREAQHAEEKGRYEAE--GQALTE---KVSLLEKA 491
Db       1176 ELK---ETHLQLDERQKKFEELEEKLKQAQQSEQKLQQESQTSKEKLTEIQQSLQELQDS 1232

Qy        492 SRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNR 551
Db       1233 VKQKEELVQNLEEKVRESSSIIEAQNTKLNESNVQL----------ENKTSCLKETQDQ 1281

Qy        552 VMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGS 611
Db       1282 LLESQKKEKQ---------------------------LQEEAAKLSGELQQVQEANG- 1311

Qy        612 SLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKE- 670
Db       1312 -------DIKDSLVKVEELVKVLEEK---LQAA---TSQLDAQQATNKELQELLVKSQEN 1358

Qy        671 ---------ALMEEIL-------KLKSLLSTKREQITTLR-------TVLKANKQTAEVA 707
Db       1359 EGNLQGESLAVTEKLQQLEQANGELKEALCQKENGLKELQGKLDESNTVLESQKKSHNEI 1418

Qy        708 LANLKSKYENEKAMVTETMMKLRNELKALK---EDAATFSSLRAMFATRCDEYITQLDEM 764
Db       1419 QDKLEQAQQKERTLQEET-SKLAEQLSQLKQANEELQKSLQQKQLLLEKGNEFDTQLAEY 1477

Qy        765 QRQLAAAEDEKKTLNSLLRMAIQQKLA---------------------LTQRLELLELD 802
Db       1478 QKVIDEMDDAASVKSALLEQ-LQNRVAELETALRQANDAQKTAYLETKELRRQLESLELE 1536


RESULT 10
B43402
myosin heavy chain-B, neuronal - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: B43402; A43402
R;Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A;Reference number: A43402; MUID:92388144; PMID:1355479
A;Accession: B43402
A;Molecule type: mRNA
A;Residues: 1-2007 <TAK>
A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48988.1; PID:g212452
A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A;Accession: A43402
A;Molecule type: mRNA
A;Residues: 1-211;222-631;653-2007 <TA2>
A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48985.1; PID:g212449
A;Note: sequence extracted from NCBI backbone (NCBIN:112864)
C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F;1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS
F;88-802/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;212-221/Region: alternatively spliced segment 1 #status experimental
F;559-593/Region: actin binding #status predicted
F;632-652/Region: alternatively spliced segment 2 #status experimental
F;692-714/Region: actin binding #status predicted
F;875-2007/Domain: coiled coil #status predicted <COI>
F;875-1315/Region: S2
F;1316-2007/Region: light meromyosin
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;732,742/Active site: Cys #status predicted
F;1954/Binding site: phosphate (Thr) (covalent) #status predicted
F;1987/Binding site: phosphate (Ser) (covalent) #status predicted


  Query Match             8.5%;  Score 347.5;  DB 1;  Length 2007;
  Best Local Similarity   21.3%;  Pred. No. 1.2e-06;
  Matches  213;  Conservative  177;  Mismatches  333;  Indels  275;  Gaps   41;

Qy          5 SEEEEYARLVME----------------AQPEWLRAEVKRLSHEL-AETTREKIQAAEYG 47
Db       1047 AEEEEKAKNLAKLKNKQEMMITDLEERLKKEEKTRQELEKAKRKLDGETTDLQDQIAELQ 1106

Qy         48 LAVLEEKHQLKLQFEELEV-----DYEAIR--SEMEQLKEAFGQAHTNHKKVAADGESRE 100
Db       1107 AQIEELKIQLAKKEEELQAALARGDEEAVQKNNALKVIRELQAQI----AELQEDLESEK 1162

Qy        101 ESLIQESASKEQYYVRKVLE-LQTELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQ 159
Db       1163 AS--RNKAEKQKRDLSEELEALKTELE---DTLDTTAAQQELRTKREQEVAELKKAIEEE 1217

Qy        160 RGRLRDDIKEYKFREARLLQDYSE-----------------LEEENISLQKQVSVLRQN 201
Db       1218 TKNHEAQIQEIRQRHATALEELSEQLEQAKRFKANLEKNKQGLESDNKELACEVKVLQQV 1277

Qy        202 QVEFEGLKHEIKRLEEETEYLNSQLEDAIRLK-EISE---------------------- 237
Db       1278 KAESE---HKRKKLDAQVQELTAKVTEGERLRVELAEKANKLQNELDNVSSLLEEAEKKG 1334

Qy        238 ----------------------------------RQLEEALETLKTEREQKNSLRKELS 262
Db       1335 IKFAKDAASLESQLQDTQELLQEETRQKLNLSSRIRQLEEEKNNLQEQQEEEEEARKNLE 1394

Qy        263 HYM------------SINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAK 310
Db       1395 KQMLALQAQLAEAKKKVDDDLGT------IEGLE--ENKKKLLKDMESLSQRLEEKAMAY 1446

Qy        311 LPLDNKTSTPKKEGLAPPSPSLVSDLLSELN-----ISEIQKLKQQLMQMEREKAGLLAT 365
Db       1447 DKLEKTKNRLQQE---------LDDLMVDLDHQRQIVSNLEKKQKKFDQMLAEEKNISAR 1497

Qy        366 LQDTQKQLEHTRGSLSEQQEKVTRLTENLSALRRLQASKE------RQTALDNEKDRDSH 419
Db       1498 YAEERDRAE------AEAREKETKALSLARALEEALEAKEEFERQNKQLRADMEDLMSSK 1551

Qy        420 EDGDYYEVDINGPEILACKYHVAVAEAGELREQLK----ALRSTHEAR-------EAQHA 468
Db       1552 DD-----VGKNVHELEKSK-RTLEQQVEEMRTQLEELEDELQATEDAKLRLEVNMQAMKA 1605

Qy        469 EEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKK-------------VSDVAG 514
Db       1606 QFERDLQARDEQNEEKKRMLVK---QVRELEAELEDERKQRALAVAAKKKMEMDLKDLEG 1662

Qy        515 ETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYRE-GQGGAGRT---SPG 570
Db       1663 QIEAA-NKARDEAI---KQLRKL------------QAQMKDYQRELEEARASRDEIFAQS 1706

Qy        571 GRTSPEARGRRSPIL-LPKGLLAPEAGRADG-----GTGDSSPSPGSSLPSPLSDPRREP 624
Db       1707 KESEKKLKGLEAEILQLQEEFAASERARRHAEQERDELADEIANSASGKSALLDEKRRLE 1766

Qy        625 MNIYNL----------IAIIRDQIKHLQAAVDR-TTELSRQRIASQELGPAVDKDKEALM 673
Db       1767 ARIAQLEEELEEEQSNMELLNERFRKTTLQVDTLNSELAGERSAAQ-------KSENARQ 1819

Qy        674 EEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENE-------KAMVTETM 726
Db       1820 QLERQNKELKAKLQELEGSVKSKFKATISTLEAKIAQLEEQLEQEAKERAAANKLVRRTE 1879

Qy        727 MKLRNELKALKEDAATFSSLRAMFATRCDEYITQLD-------EMQRQLAAAEDEKKTLN 779
Db       1880 KKLKEVFMQVEDE-----------RRHADQYKEQMEKANARMKQLKRQLEEAEEEATRAN 1928

Qy        780 SLLRMAIQQKL--------ALTQRLELLELDHEQTRRG 809
Db       1929 A-SRRKLQRELDDATEANEGLSREVSTLK---NRLRRG 1962
```

```
RESULT 11
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
A;Cross-references: GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
F;1-818/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

  Query Match             8.5%;  Score 346.5;  DB 2;  Length 2116;
  Best Local Similarity   21.0%;  Pred. No. 1.5e-06;
  Matches  224;  Conservative  161;  Mismatches  338;  Indels  345;  Gaps   43;

Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEK--HQLKLQFEE 63
              :||:  |  :| : : | | :: :  :|  |  ||   : | | | :|     |: |  |
Db       1104 DEEKKNRDALEKKKKALDAMLEEMKDQLESTGGEK--KSLYDLKVKQESDMEALRNQISE 1161

Qy         64 --------------LEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESAS 109
                           || :   :: |:|  : |        |||  | | :   | :|:|:
Db       1162 LQSTIAKLEKIKSTLEGEVARLQGELEAEQLAKSNVEKQKKKVELDLEDKSAQLAEETAA 1221

Qy        110 KE------QYYVRKVLELQTELKQLRNVLTNTQSENERLAS------------------- 144
              |:      :   ::: |:||:| :  |   |: | |: | :
Db       1222 KQALDKLKKKLEQELSEVQTQLSEANNKNVNSDSTNKHLETSFNNLKLELEAEQKAKQAL 1281

Qy        145 ------VAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELE---EENISLQKQV 195
                   :  ||| :|: :| .:: :   : |    |:  | :: |||:   || :: :| |
Db       1282 EKKRLGLESELKHVNEQLEEEKKQKESNEK----RKVDLEKEVSELKDQIEEEVASKKAV 1337

Qy        196 SVLR-QNQVEFEGLK--------------HEIKRLEEETEYLNS-------QLEDAIRLK 233
              :  : : : | : :|               ::| |: : | | :       ||: | | |
Db       1338 TEAKNKKESELDEIKRQYADVVSSRDKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSK 1397

Qy        234 EISERQLEEALETLKTEREQK--------------NSLRKELSHYMSINDSFYT--SHLH 277
              : :|  ||||:: |: |  :|               | : ||    :::   |     |:
Db       1398 KKAEFDLEEAVKNLEEETAKKVKAEKAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLN 1457

Qy        278 VSLDGLK-FSDDAAEPNNDA-----------EALVNGFEHGGLAKLPLDNKTSTPKKEGL 325
              |  |:   ::| |  | |             |:| :  :    ||   : | |   :  :
Db       1458 EELSELRSVLEEADERCNSAIKAKKTAESALESLKDEIDAANNAKAKAERK-SKELEVRV 1516

Qy        326 APPSPSLVSDLLSELNISEIQKLKQQL----MQMEREKAGLLATLQD---TQKQLEHTRG 378
              |   ||  |    :|:  |:|   ::     :::||    : : :|   |:||
Db       1517 AELEESL-EDKSGTVNVEFIRKKDAEIDDLRARLDRETESRIKSDEDKKNTRKQFADLEA 1575

Qy        379 SLSEQQEKVT---RLTENLSA-------------LRRLQASKERQTALDNEKDRDSHEDG 422
               : | | :|    || : | :                |::  | ::        :| : |:|
Db       1576 KVEEAQREVVTIDRLKKKLESDIIDLSTQLDTETKSRIKIEKSKKKLEQTLAERRAAEEG 1635

Qy        423 D------------YYEVD------------INGPEILACKYHVAVAEAGELREQL----- 453
                         : |||             :|  |    |    |||  |::|||
Db       1636 SSKAADEEIRKQVWQEVDELRAQLDSERAALNASE---KKIKSLVAEVDEVKEQLEDEIL 1692

Qy        454 ------KALRS----THEAREAQHAEE--------------------KGRYEAE------ 477
                   || |:       | |:    ||                      | :|:||
Db       1693 AKDKLVKAKRALEVELEEVRDQLEEEEDSRSELEDSKRRLTTEVEDIKKKYDAEVEQNTK 1752

Qy        478 ----GQALTEKVSLLEK-----------------ASRQDRELLARLEKELKKVS----- 510
                 : ||: |  |:|                    :: : ||:|: |:|  |
Db       1753 LDEAKKKLTDDVDTLKKQLEDEKKKLNESERAKKRLESENEDFLAKLDAEVKNRSRAEKD 1812

Qy        511 -----------------DVAGETQGSLSVAQ-----DEL-----------------VTF 530
                              : | :||  :  |:     |||                   |
Db       1813 RKKYEKDLKDTKYKLNDEAATKTQTEIGAAKLEDQIDELRSKLEQEQAKATQADKSKKTL 1872

Qy        531 SEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGL 590
              |: ||   :     |   :: :   :| :   |       |  ||   :|     | |
Db       1873 EGEIDNLRAQI-----EDEGKIKMRLEKEKRALEGELEELRETVEEAEDSKSEAEQSKRL 1927

Qy        591 LAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQ-AAVDRTT 649
              :  |                  | | ||     :   |  |   :||   |:
Db       1928 VELE---------------------LEDARRNLQKEIDAKEIAEDAKSNLQREIVEAKG 1965

Qy        650 ELSRQRIASQELGPAVDKDKEALMEEILKLKSLL----STKREQITTLRTVLKANKQTAE 705
               |  : ||        |: :: |  ||  | : :      | :||       | ||:  |
Db       1966 RLEEESIARTN----SDRSRKRLEAEIDALTAQVDAEQKAKNQQI-------KENKK-IE 2013

Qy        706 VALANLKSKY-ENEKAMVTETMM--KLRNELKALKEDAATFSSLRAMFATRCDEYITQLD 762
               |   : |: |:||    | ::  ||  : |  |::||                   |
Db       2014 TELKEYRKKFGESEKTKTKEFLVVEKLETDYKRAKKEAA-------------------D 2053

Qy        763 EMQRQLAAAEDEKKTLN--SLLRMAIQQKLALTQRLELLELDHEQTRR 808
              | |::|    | :| |:  |||: || :          |: ||::|:|
Db       2054 EQQQRLTVENDLRKHLSEISLLKDAIDK----------LQRDHDKTKR 2091

RESULT 12
A59252
myosin heavy chain, nonmuscle, form IIB - human
N;Alternate names: myosin type 10; NMMHC-B
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C;Accession: A59252; B61231; G02055
R;Phillips, C.L.; Yamakawa, K.; Adelstein, R.S.
J. Muscle Res. Cell. Motil. 16, 379-389, 1995
A;Title: Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analys
A;Reference number: A59252; MUID:96025307; PMID:7499478
A;Accession: A59252
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1976 <SIM>
A;Cross-references: GB:M69181; NID:g641957; PIDN:AAA99177.1; PID:g641958
A;Experimental source: clone lib Lambda Zap II adult human T-cell library; cell line
A;Note: between nucleotides 1942-1943 in mRNA encoding human brain MHC-B there is an
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelste
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on diff
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: B61231
A;Molecule type: mRNA
A;Residues: 63-237,'K',239-664,'L',666-722 <SI2>
A;Cross-references: GB:M69181; NID:g641957
R;Weir, L.
submitted to the EMBL Data Library, August 1995
A;Reference number: H00753
A;Accession: G02055
```

```
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-81 <WEI>
A;Cross-references: EMBL:U34304; NID:g1143217; PIDN:AAA84880.1; PID:g1143218
C;Genetics:
A;Gene: GDB:MYH10
A;Cross-references: GDB:127350; GDB:G00-127-350; OMIM:160776
A;Map position: 17p13-17p13
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F;88-771/Domain: myosin motor domain homology <MMO>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;559-572/Region: actin binding #status predicted
F;633-647/Region: actin binding #status predicted
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;701,711/Active site: Cys #status predicted

  Query Match             8.4%;  Score 346;  DB 2;  Length 1976;
  Best Local Similarity   21.8%;  Pred. No. 1.4e-06;
  Matches  207;  Conservative  170;  Mismatches  306;  Indels  268;  Gaps   39;

Qy          5 SEEEEYAR---LVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQF 61
              :|||| |:    :  |  :   :||  |  | ||:::: |:  |   |   |:  |
Db       1016 AEEEEKAKNLAKIRNKQEVMISDLEERLKKE--EKTRQELEKAKRKLD--GETTDLQDQI 1071

Qy         62 EELEVDYEAIRSEMEQLKE------AFGQAHTNHK----KVAADGESR----EESLIQES 107
               ||:   : :: :: : :|      | |   | ||    ||  : :::    :|    |
Db       1072 AELQAQIDELKLQLAKKEEELQGALARGDDETLHKNNALKVVRELQAQIAELQEDFESEK 1131

Qy        108 ASKEQYYVRKVLELQTELK----QLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRL 163
              ||: :   :|  :|  ||:    :| : |  |. :: |      ||: |: : :| :
Db       1132 ASRNKAEKQK-RDLSEELEALKTELEDTLDTTAAQQELRTKREQEVAELKKALEEETKNH 1190

Qy        164 RDDIKEYKFREARLLQDYSE-----------------LEEENISLQKQVSVLRQNQVEF 205
              |:: : | |  |:: ||                  || :|  |  :| ||:| : |
Db       1191 EAQIQDMRQRHATALEELSEQLEQAKRFKANLEKNKQGLETDNKELACEVKVLQQVKAES 1250

Qy        206 EGLKHEIKRLEEETEYLNSQLEDAIRLK-EISER--QLEEALE---TLKTEREQKNSLRK 259
              |  |: |:|: : : |:::: :  ||: |::|:  :|:  |:   ||  | |:|
Db       1251 E---HKRKKLDAQVQELHAKVSEGDRLRVELAEKASKLQNELDNVSTLLEEAEKK----- 1302

Qy        260 ELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN---DAEALVNGFEHGGLAKLPLDNK 316
                                 |:||: |||   :   | : |:                :
Db       1303 ----------------------GIKFAKDAASLESQLQDTQELL--------------Q 1325

Qy        317 TSTPKKEGLAPPSPSLVSDLLSELN-ISEIQKLKQQLMQ-MEREKAGLLATLQDTQKQLE 374
              | :|  |:    | :  |  | | : | |: ::  : :|::   | : | |:|:|:::
Db       1326 EETRQKLNLS----SRIRQLEEEKNSLQEQQEEEEEARKNLEKQVLALQSQLADTKKKVD 1381

Qy        375 HTRGSLSEQQEKVTRLTENLSAL-RRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPE 433
                |::   :|   :| :: || :||:       |:  |:|   |  |   ||::
Db       1382 DDLGTIESLEEAKKKLLKDAEALSQRLEEKALAYDKLEKTKNRLQQELDD-LTVDLDHQR 1440

Qy        434 ILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAE----------GQALT 482
              :|       :  :|  : |::       |::|||: | |||           :||
Db       1441 QVASNLEKKQKKFDQLLAEEKSI-------SARYAEERDRAEAEAREKETKALSLARALE 1493

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVA------GETQGSLSVAQDELVTFSEEL-- 534
              | :   |:  ||:::| | :|  :    ||       ::: :|    :|: |  |||
Db       1494 EALEAKEEFERQNKQLRADMEDLMSSKDDVGKNVHELEKSKRALEQQVEEMRTQLEELED 1553

Qy        535 -------ANLYHHVCMC---------------NNETPNRVMLDYYREGQGGAGRTSPGGR 572
                    | |   | |                  ||   |:::   ||
Db       1554 ELQATEDAKLRLEVNMQAMKAQFERDLQTRDEQNEEKKRLLIKQVRE------------- 1600

Qy        573 TSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE--------- 623
                                                  | : | | |::
Db       1601 ----------------------------------------LEAELEDERKQRALAVASKK 1620

Qy        624 --PMNIYNLIAII------RD----QIKHLQAAV-DRTTELSRQRIASQEL---GPAVDK 667
                ::: :| | |      ||    |:: ||| : |   ||   | :  |:       :|
Db       1621 KMEIDLKDLEAQIEAANKARDEVIKQLRKLQAQMKDYQRELEEARASRDEIFAQSKESEK 1680

Qy        668 DKEALMEEILKL-KSLLSTKR------------EQITTL---RTVLKANKQTAEVALAN 710
                ::|  |||:| : | |::|            ::||     :: |   |:  |  :|
Db       1681 KLKSLEAEILQLQEELASSERARRHAEQERDELADEITNSASGKSALLDEKRRLEARIAQ 1740

Qy        711 LKSKYENEKA---MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQ 767
              |: : | |::   :: :   |   ::  |  : |   |     | : |    ||:   ::
Db       1741 LEEELEEEQSNMELLNDRFRKTTLQVDTLNAELAAERSA----AQKSDNARQQLERQNKE 1796

Qy        768 LAAAEDEKKTLNSLLRMAIQQKL-----ALTQRLELLELDHEQTRRGRAKA 813
              | |   |       |  |:: |      ||  ::  ||   ||  : || |
Db       1797 LKAKLQE-------LEGAVKSKFKATISALEAKIGQLEEQLEQEAKERAAA 1840


RESULT 13
A41604
myosin heavy chain, smooth muscle, long splice form - rabbit
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: A41604; A33501
R;Babij, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A;Title: Characterization of a mammalian smooth muscle myosin heavy-chain gene: compl
A;Reference number: A41604; MUID:92073350; PMID:1961735
A;Accession: A41604
A;Molecule type: mRNA
A;Residues: 1-1972 <BAB>
A;Cross-references: GB:M77812
R;Nagai, R.; Kuro-o, M.; Babij, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A;Title: Identification of two types of smooth muscle myosin heavy chain isoforms by
A;Reference number: A33501; MUID:89255535; PMID:2722872
A;Accession: A33501
A;Molecule type: mRNA
A;Residues: 1455-1972 <NAG>
A;Cross-references: GB:J04833; NID:g165519; PIDN:AAA31407.1; PID:g165520
A;Experimental source: smooth muscle
A;Note: examination by Southern blotting for the regions of difference between this i
ative splicing
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methyla
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;559-572/Region: actin binding #status predicted
F;633-647/Region: actin binding #status predicted
F;844-1938/Domain: coiled coil #status predicted <COI>
F;844-1284/Region: S2
F;1285-1972/Region: light meromyosin
F;1939-1972/Domain: carboxyl-terminal <CBT>
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;701,711/Active site: Cys #status predicted

  Query Match             8.4%;  Score 344.5;  DB 1;  Length 1972;
  Best Local Similarity   21.8%;  Pred. No. 1.6e-06;
  Matches  200;  Conservative  157;  Mismatches  312;  Indels  249;  Gaps   37;

Qy         20 EWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQ 77
              :| |   |     | : ||  |::|| |  |  ::|: |                ||:::
Db       833 QWWRLFTK--VKPLLQVTRQEEEMQAKEDELQKIKERQQ-------------KAESELQE 877

Qy         78 LKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQS 137
              |:    | ||          | |::|:||    |         ||  | :::|
Db       878 LQ----QKHTQ---------LSEEKNLLQEQLQAE-------TELYAEAEEMR-------- 910

Qy        138 ENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSV 197
                 |||:  |||:||   :| :     |  :: :    :: |   :|||:   |::: :
Db       911 --VRLAAKKQELEEILHEMEARLEEEEDRGQQLQAERKKMAQQMLDLEEQ---LEEEEAA 965

Qy        198 LRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSL 257
              :: |:|    : :|||:||::    :: |     : ::: | :: :    |  | |:  :|
Db       966 RQKLQLEKVTAEAKIKKLEDDILVMDDQNNKLSKERKLLEERISDLTTNLAEEEEKAKNL 1025
```

```
Qy        258 RKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKT 317
Db       1026 TK----------------------LK---------NKHESMISELE----VRLKKEEK- 1048

Qy        318 STPKKEGLAPPSPSLVSDL---LSELNISEIQKLKQQLMQMEREKAGLLATLQD------ 368
Db       1049 SRQELEKLKRKMDGEASDLHEQIADLQ-AQIAELKMQLAKKEEELQAALARLEDETSQKN 1107

Qy        369 -TQKQLEHTRGSLSEQQEKVTR--------------LTENLSALR--------------R 399
Db       1108 NALKKIRELEGHISDLQEDLDSERAARNKAEKQKRDLGEELEALKTELEDTLDTTATQQE 1167

Qy        400 LQASKERQT-----ALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLK 454
Db       1168 LRAKREQEVTVLKKALDEE--TRSHE-AQVQEMR---------QKHTQVVE--ELTEQLE 1213

Qy        455 A-------LRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDREL--------- 498
Db       1214 QFKRAKANLDKTKQTLEKENADLAGELRVLGQAKQEVEHKKKKLEVQLQELQSKCSDGER 1273

Qy        499 -LARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVC----MCNNETPNRVM 553
Db       1274 ARAELNDKVHKLQNEVESVTGMLSEAEGKAIKLAKEVASLGSQLQDTQELLQEETRQKLN 1333

Qy        554 LDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSL 613
Db       1334 VS-TKLRQLEDERNSLQEQLDEEMEAKQN---LERHI-------------------STL 1369

Qy        614 PSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALM 673
Db       1370 NIQLSDSKKKLQDFASTVESLEEGKKRFQKEIESLTQQYEEKAAAYD---KLEKTKNRLQ 1426

Qy        674 EEILKLKSLLSTKREQITTLRTVLKANKQTAEVAL--ANLKSKYENEK------------ 719
Db       1427 QELDDLVVDLDNQRQLVSNLE---KKQKKFDQLLAEEKNISSKYADERDRAEAEAREKET 1483

Qy        720 -----AMVTETMMKLRNE-------LKALKEDAATF------------SSLRAMFATRCD 755
Db       1484 KALSLARALEEALEAKEELERTNKMLKAEMEDLVSSKDDVGKNVHELEKSKRAL-ETQME 1542

Qy        756 EYITQLDEMQRQLAAAEDEKKTLN---SLLRMAIQQKL------------ALTQRLELLE 800
Db       1543 EMKTQLEELEDELQATEDAKLRLEVNMQALKVQFERDLQARDEQNEEKRRQLQRQLHEYE 1602

Qy        801 LDHEQTRRGRAKAAPKTK 818
Db       1603 TELEDERKQRALAAAAKK 1620
```

RESULT 14
S03166
myosin heavy chain, gizzard smooth muscle [similarity] - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: S03166; A27066; A26045; A36604; A43298
R;Yanagisawa, M.; Hamada, Y.; Katsuragawa, Y.; Imamura, M.; Mikawa, T.; Masaki, T.
J. Mol. Biol. 198, 143-157, 1987
A;Title: Complete primary structure of vertebrate smooth muscle myosin heavy chain deduc
A;Reference number: S03166; MUID:88118918; PMID:2892941
A;Accession: S03166
A;Molecule type: mRNA
A;Residues: 1-1979 <YAN>
A;Cross-references: EMBL:X06546; NID:g63633; PIDN:CAA29793.1; PID:g63634
A;Note: part of this sequence was confirmed by protein sequencing
R;Maita, T.; Onishi, H.; Yajima, E.; Matsuda, G.
J. Biochem. 102, 133-145, 1987
A;Title: Amino acid sequence of the amino-terminal 24 kDa fragment of the heavy chain of
A;Reference number: A27066; MUID:88032919; PMID:3312184
A;Accession: A27066
A;Molecule type: protein
A;Residues: 2,'2',4-204 <MAI>
R;Onishi, H.; Maita, T.; Miyanishi, T.; Watanabe, S.; Matsuda, G.
J. Biochem. 100, 1433-1447, 1986
A;Title: Amino acid sequence of the 203-residue fragment of the heavy chain of chicke
A;Reference number: A26045; MUID:87194651; PMID:3571180
A;Accession: A26045
A;Molecule type: protein
A;Residues: 653-855 <ONI>
R;Onishi, H.; Maita, T.; Matsuda, G.; Fujiwara, K.
J. Biol. Chem. 265, 19362-19368, 1990
A;Title: Lys-65 and Glu-168 are the residues for carbodiimide-catalyzed cross-linking
A;Reference number: A36604; MUID:91035476; PMID:1977747
A;Accession: A36604
A;Status: preliminary
A;Molecule type: protein
A;Residues: 54-67;146-183 <ON2>
R;Cole, D.G.; Yount, R.G.
Biochemistry 31, 6186-6192, 1992
A;Title: Stability and photochemical properties of vanadate-trapped nucleotide comple
A;Reference number: A43298; MUID:92329440; PMID:1385724
A;Accession: A43298
A;Status: preliminary
A;Molecule type: protein
A;Residues: 169-183 <COL>
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; blocked amino end; coiled coil; hydrolase; methylated
F;87-777/Domain: myosin motor domain homology <MMOT>
F;177-184/Region: nucleotide-binding motif A (P-loop)
F;565-578/Region: actin binding #status predicted
F;639-653/Region: actin binding #status predicted
F;850-1940/Domain: coiled coil <COI>
F;850-1290/Region: S2
F;1291-1979/Region: light meromyosin
F;1941-1979/Domain: carboxyl-terminal <CBT>
F;2/Modified site: blocked amino end (Ser) (in mature form) #status experimental
F;128/Modified site: N6,N6,N6-trimethyllysine (Lys) #status experimental
F;183/Binding site: ATP (Lys) #status predicted
F;707,717/Active site: Cys #status predicted

Query Match 8.3%; Score 338.5; DB 1; Length 1979;
Best Local Similarity 20.4%; Pred. No. 2.8e-06;
Matches 196; Conservative 147; Mismatches 335; Indels 281; Gaps 31;

```
Qy         20 EWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEE--------KHQLKLQFEELEVDYE 69
Db       1096 EELQAALARLEDETSQKNNALKKIRELESHISDLQEDLESEKAARNKAEKQKRDLSEELE 1155

Qy         70 AIRSEME----------QLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVL 119
Db       1156 ALKTELEDTLDTTATQQELRAKREQEVTVLKRALEEETRTHEAQVQEMRQK---HTQAVE 1212

Qy        120 ELQTELKQLRNVLTNTQSENERL----ASVAQELKEINQ----------NVEIQ------ 159
Db       1213 ELTEQLEQFKRAKANLDKTKQTLEKDNADLANEIRSLSQAKQDVEHKKKKLEVQLQDLQS 1272

Qy        160 --------RGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVEFEGLKHE 211
Db       1273 KYSDGERVRTELNEKVHKLQIEVENVTSLLNEAESKNIKLTKDVATLGSQLQDTQELLQE 1332

Qy        212 -----------IKRLEEETEYLNSQLEDAIRLKEISER-------------QLEEALET 246
Db       1333 ETRQKLNVTTKLRQLEDDKNSLQEQLDEEVEAKQNLERHISTLTIQLSDSKKKLQEFTAT 1392

Qy        247 LKTEREQKNSLRKELSHYMSINDSF-----------YTSHLHVSLDGL---------- 283
Db       1393 VETMEEGKKKLQREIE---SLTQQFEEKAASYDKLEKTKNRLQQELDDLVVDLDNQRQLV 1449

Qy        284 --------KFSDDAAEPNN-----------DAEA---------LVNGFEHGGLAKLPLD 314
Db       1450 SNLEKKQKKFDQMLAEEKNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELE 1509

Qy        315 NKTSTPKKEGLAPPSPSLVSDLLSE-----LNISEIQKLKQQLMQMEREKAGLLATLQDT 369
Db       1510 RTNKMLKAE---------MEDLVSSKDDVGKNVHELEKSKRTLEQQ----------VEEM 1550
```

```
Qy        370 QKQLEHTRGSLSEQQEKVTRLTENLSAL-----RRLQASKERQTALDNEKDRDSHEDGDY 424
Db       1551 KTQLEELEDELQAAEDAKLRLEVNMQAMKSQFERDLQARDEQ-----NEEKR-------- 1597

Qy        425 YEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEA-------REAQHAEEKGRYEAE 477
Db       1598 -------------------------RQLLKQLHE-HETELEDERKQRALAAAAKKKLEVD 1631

Qy        478 GQALTEKVSLLEKASRQDRELLARLEKELK--------------KVSDVAGETQGSLSVA 523
Db       1632 VKDLESQVDSANKAREEAIKQLRKLQAQMKDYQRDLDDARAAREEIFATARENEKKAKNL 1691

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
Db       1692 EAELIQLQEDLAAAER----------ARKQADLEKEEMAEELASANSGRTSLQDEKRR-- 1739

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQI-KHLQ 642
Db       1740 ----------------------------LEARIAQLEEELDEEHSNIETMSDRMRKAVQ 1770

Qy        643 AAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQ 702
Db       1771 QAEQLNNELATERATAQ-------KNENARQQLERQNKELRSKLQEMEGAVKSKFKSTIA 1823

Qy        703 TAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLR---AMFATRCDEYIT 759
Db       1824 ALEAKIASLEEQLEQEAREKQAAAKTLRQKDKKLKDALLQVEDERKQAEQYKDQAEKGNL 1883

Qy        760 QLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
Db       1884 RLKQLKRQLEEAEEESQRIN-----ANRRKL---QR-ELDEATESNDALGREVAALKSK 1933


RESULT 15
A59236
embryonic muscle myosin heavy chain - sea squirt (Halocynthia roretzi)
C;Species: Halocynthia roretzi
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C;Accession: A59236
R;Araki, I.
submitted to GenBank, February 1999
A;Reference number: A59236
A;Accession: A59236
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1927 <ARA>
A;Cross-references: GB:D45163; NID:g1197167; PIDN:BAA08111.1; PID:g1197168
A;Experimental source: clone lib lambda gt11; dev stage tailbud embryo
C;Genetics:
A;Gene: MHCemb
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;89-766/Domain: myosin motor domain homology <MMO>

  Query Match             8.2%;  Score 337.5;  DB 2;  Length 1927;
  Best Local Similarity   21.2%;  Pred. No. 3e-06;
  Matches  201;  Conservative  166;  Mismatches  370;  Indels  211;  Gaps   37;

Qy          5 SEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAE-YGLAVLEEKHQLKLQFEE 63
Db        984 SQDEQIAKLQKE----------KKALQEAHQQTLDDLQSEEDKANSLTKQKAKLEQQVDD 1033

Qy         64 LEVDYE---AIRSEMEQLKEAFGQAHTNHKKVAADGESREESL---------IQESASKE 111
Db       1034 LEASLEQEKKLRMDLERTK----------RKLEGDLRLTQETVMDLENDKQRLEEKLKKQ 1083

Qy        112 QYYVRK-----------VLELQTELKQLRNVLTNTQSENE-----------RLASVAQEL 149
Db       1084 EFEYSQLATKLEDEQALVMQLQKKIKELQARIEELEEELEAERAARAKVEKQRADLSREL 1143

Qy        150 KEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQV------ 203
Db       1144 EELSERLEEAGGATAAQIELNKRREAEFSKLRRELEESNLAHEATVSTLRKKHADSSAEM 1203

Qy        204 --EFEGLKHEIKRLEEETEYLNSQLED-AIRLKEISE---------RQLEEALETLKTER 251
Db       1204 SEQIDNLQRVKQKLEKEKSEMKMEVDDLAANVESITKAKLNYEKMARNLEEQFSESKTKC 1263

Qy        252 E----QKNSLRKELSHYMSINDSF-----YTSHLHVSLDGLKFSDDAAEPNNDAEALVNG 302
Db       1264 DNFCKEVNELNAAKARFASENGELSRQLEEREHLMAQLTRTKNSS-----SQQIEELKRV 1318

Qy        303 FEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLM--------- 353
Db       1319 VEEETKAKAALAHSVQASRHD----------NDLLRE-QYEEEQEAKAELQRALSKANAE 1367

Qy        354 -----------------QMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSA 396
Db       1368 VAQWRNKYETDAIQRTEELEEAKKKLATRLQEAEEQVEATQAKCASLDKTKNRLQGELED 1427

Qy        397 LR-RLQASKERQTALDNEK--------DRDSHEDGDYYEVDINGPEILACKYHVAVAEAG 447
Db       1428 LTIDLERSNSAAAALDKKQRNFDKVLAERKQKEE----EIQV--------ELEQAQKEAR 1475

Qy        448 ELREQLKALRSTHE----AREAQHAEEKGRYEAEGQALTEK-------VSLLEKASRQDR 496
Db       1476 GLSTELFKMKNSYEESLDALETVKRENKNLQE-EIADLTDQLGEGGKSIHELEKAKR--- 1531

Qy        497 ELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDY 556
Db       1532 ----TLEHERNEIQAALEEAEGAIEGEESKVLRLQVELAQIKQDFERRLSEKEEEI--EN 1585

Qy        557 YREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSP 616
Db       1586 TRRNQQRA-LESMQTTLDSESKSRQEAVRMKK--------KMEGDLND--------LEIQ 1628

Qy        617 LSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEI 676
Db       1629 LGHATRQASESQKSVKTFQAHVKDLELQVD---ESQRHSDDLQEQFAVIERRENLIKAEI 1685

Qy        677 LKLKSLL-----STKREQITTLRTVLKAN-KQTAEVALANLKSKYENE----KAMVTETM 726
Db       1686 DELRSALEQAERGRKLAETELLESSERSNLLHTQNTALINQKRKLEGELQNVQSEVEEAV 1745

Qy        727 MKLRN-ELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQL-AAAEDEKKTLNSLLRM 784
Db       1746 QEQRNAEEKAKK--AITDAAMMAEELKKEQDLSSHLERMKKNTEQTVKDLQQRLDEAEQV 1803

Qy        785 AIQQKLALTQRLELL------ELDHEQTRRGRA-----KAAPKTKPAT 821
Db       1804 ALKGGKKQVQKLETRVRELENELDSEQRRNGESVKNQRKTERKLKEVT 1851


Search completed: July  1, 2003, 08:32:22
Job time : 34.792 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 21.1764 Seconds
(without alignments)
1360.244 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5183	100.0	979	3	US-08-870-529-2	Sequence 2, Appli
2	502.5	9.7	345	3	US-09-173-581-6	Sequence 6, Appli
3	502.5	9.7	345	4	US-09-420-915-6	Sequence 6, Appli
4	476	9.2	699	4	US-09-457-040B-18	Sequence 18, Appl
5	441	8.5	302	3	US-09-221-235-2	Sequence 2, Appli
6	441	8.5	302	3	US-09-221-928-2	Sequence 2, Appli
7	441	8.5	302	4	US-09-221-527-2	Sequence 2, Appli
8	441	8.5	302	4	US-09-221-236-2	Sequence 2, Appli
9	441	8.5	302	4	US-09-221-416-2	Sequence 2, Appli
10	441	8.5	302	4	US-09-221-245-2	Sequence 2, Appli
11	441	8.5	302	4	US-09-163-115-2	Sequence 2, Appli
12	441	8.5	302	4	US-09-221-528-2	Sequence 2, Appli
13	441	8.5	302	4	US-09-593-553-2	Sequence 2, Appli
14	441	8.5	302	4	US-09-221-237-2	Sequence 2, Appli
15	403.5	7.8	233	2	US-08-712-709-1	Sequence 1, Appli
16	403.5	7.8	233	3	US-09-111-444-1	Sequence 1, Appli
17	403.5	7.8	233	4	US-09-541-228-1	Sequence 1, Appli
18	391.5	7.6	293	2	US-09-049-671-1	Sequence 1, Appli
19	391.5	7.6	293	3	US-09-295-068-1	Sequence 1, Appli
20	385	7.4	1037	4	US-09-428-711A-21	Sequence 21, Appl
21	368.5	7.1	239	2	US-08-712-709-7	Sequence 7, Appli
22	368.5	7.1	239	2	US-09-049-671-3	Sequence 3, Appli
23	368.5	7.1	239	3	US-09-295-068-3	Sequence 3, Appli
24	368.5	7.1	239	3	US-09-111-444-7	Sequence 7, Appli
25	368.5	7.1	239	4	US-09-541-228-7	Sequence 7, Appli
26	357.5	6.9	416	1	US-08-252-995D-2	Sequence 2, Appli
27	357.5	6.9	416	2	US-08-834-108-2	Sequence 2, Appli
28	357.5	6.9	464	1	US-08-252-995D-6	Sequence 6, Appli
29	357.5	6.9	464	2	US-08-834-108-6	Sequence 6, Appli
30	357.5	6.9	925	1	US-08-252-995D-4	Sequence 4, Appli
31	357.5	6.9	925	2	US-08-834-108-4	Sequence 4, Appli
32	354	6.8	603	4	US-09-198-122-2	Sequence 2, Appli
33	354	6.8	685	2	US-08-878-989-1	Sequence 1, Appli
34	354	6.8	685	3	US-09-136-282-2	Sequence 2, Appli
35	354	6.8	685	4	US-09-272-796-1	Sequence 1, Appli
36	354	6.8	685	4	US-09-505-744-2	Sequence 2, Appli
37	351.5	6.8	273	1	US-08-252-995D-10	Sequence 10, Appl
38	351.5	6.8	273	2	US-08-834-108-10	Sequence 10, Appl
39	349	6.7	272	1	US-08-252-995D-12	Sequence 12, Appl
40	349	6.7	272	2	US-08-834-108-12	Sequence 12, Appl
41	348	6.7	607	2	US-08-878-989-15	Sequence 15, Appl
42	348	6.7	607	4	US-09-272-796-15	Sequence 15, Appl
43	347	6.7	591	3	US-09-082-737-2	Sequence 2, Appli
44	347	6.7	603	4	US-09-311-311C-26	Sequence 26, Appl
45	346.5	6.7	270	2	US-08-852-743-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-870-529-2
; Sequence 2, Application US/08870529
; Patent No. 6080557
; GENERAL INFORMATION:
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Bird, Timothy A.
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: IL-1/TNF-(-ACTIVATED KINASE (ITAK),
; TITLE OF INVENTION: AND METHODS OF MAKING AND USING THE SAME
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,529
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 480052.418
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 979 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-870-529-2

Query Match 100.0%; Score 5183; DB 3; Length 979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
```

```
Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979
              ||||||||||||||||||
Db        961 DLDSDSWCLLGTDSCRPSL 979


RESULT 2
US-09-173-581-6
; Sequence 6, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/173,581A
;  CURRENT FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 6
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 1567782
US-09-173-581-6

  Query Match             9.7%;  Score 502.5;  DB 3;  Length 345;
  Best Local Similarity   33.8%;  Pred. No. 1.9e-34;
  Matches  112;  Conservative   63;  Mismatches  101;  Indels   55;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   |                                  | :: |:: ||  |:  :|
Db         64 VQYRESF-------------------------------EGILDWFV-QICLALKHVHDRK 91

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db         92 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 151

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| ||| |:| :  || :||: |  :    |  || :|  :|     :
Db        152 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 209

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :        ||       |      | ||| |  ::::
Db        210 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 269

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::       |  | |   :|
Db        270 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 300


RESULT 3
US-09-420-915-6
; Sequence 6, Application US/09420915
; Patent No. 6264947
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/420,915
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/173,581
;  EARLIER FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 6
```

```
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 1567782
US-09-420-915-6

  Query Match             9.7%;  Score 502.5;  DB 4;  Length 345;
  Best Local Similarity   33.8%;  Pred. No. 1.9e-34;
  Matches  112;  Conservative   63;  Mismatches  101;  Indels   55;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   |                                | :: |:: ||  |:  :|
Db         64 VQYRESF------------------------------EGILDWFV-QICLALKHVHDRK 91

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: |||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db         92 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 151

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| :  || :||: |  :    |  || :|  :|     :
Db        152 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 209

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :         ||       |      | ||| |  ::::
Db        210 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 269

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |  .:|   ::      |   | |   :|
Db        270 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 300


RESULT 4
US-09-457-040B-18
; Sequence 18, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
;  APPLICANT: Vertex Pharmaceuticals Incorporated
;  APPLICANT:  Bellon, Steve
;  TITLE OF INVENTION: Crystallized P38 Complexes
;  FILE REFERENCE: VPI/98-14
;  CURRENT APPLICATION NUMBER: US/09/457,040B
;  CURRENT FILING DATE:  1999-12-08
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 699
;   TYPE: PRT
;   ORGANISM: EMENI - Aspergillus Nidulans
US-09-457-040B-18

  Query Match             9.2%;  Score 476;  DB 4;  Length 699;
  Best Local Similarity   35.6%;  Pred. No. 1.1e-31;
  Matches  113;  Conservative   59;  Mismatches  109;  Indels   36;  Gaps    6;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :  :| |:||     :|  |  ::  ||::  ::| |||     |  ||: |:| ||
Db         11 YEVLEKIGCGSFGIIRKVKRKSDGFILCRKEINYIKMSTKEREQLTAEFNILSSLRHPNI 70

Qy        112 IAYYN--HFMDNTTLLIELEYCNGGNL--YDKILRQKDKLFEEEMVVWYLFQIVSAVSCI 167
              :|||:  |   :  | : :||| ||:|    | |:: :|  ||: |   | |:|:|:
Db         71 VAYYHREHLKASQDLYLYMEYCGGGDLSMVIKNLKRTNKYAEEDFVWRILSQLVTALYRC 130

Qy        168 H--------------------------KAGILHRDIKTLNIFLTKANLIKLGDYGLAKK 200
              |                          :  |||||:|  ||||   | :||||:||:|
Db        131 HYGTDPAEVGSNLLGPAPKPSGLKGKQAQMTILHRDLKPENIFLGSDNTVKLGDFGLSKL 190

Qy        201 LNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLC 260
              ::| :  | | ||||:||||||:|   ||  :|||||||||:::||   :  |:|  : |

Db        191 MHS-HDFASTYVGTPFYMSPEICAAEKYTLRSDIWAVGCIMYELCQREPPFNARTHIQLV 249

Qy        261 VKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVT 320
              || :|  |   |   || ||  :: |||  :|: ||    |:: |::|  |||:|  :
Db        250 QKIREGKFAPLPDF--YSSELKNVIASCLRVNPDHRPDTATLINTPVIRLMRREVE--LN 305

Qy        321 LLNAPTKRPRSSTVTEA 337
              |:   ::   :|: :|
Db        306 NLSRAARKREEATMQKA 322


RESULT 5
US-09-221-235-2
; Sequence 2, Application US/09221235
; Patent No. 6043040
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,235
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-235-2

  Query Match             8.5%;  Score 441;  DB 3;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||      |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |     |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 6
US-09-221-928-2
; Sequence 2, Application US/09221928
; Patent No. 6121030
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,928
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-221-928-2

  Query Match             8.5%;  Score 441;  DB 3;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | |.|:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 7
US-09-221-527-2
; Sequence 2, Application US/09221527
; Patent No. 6146832
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,527
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-527-2

  Query Match             8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | |.|:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 8
US-09-221-236-2
; Sequence 2, Application US/09221236
; Patent No. 6146841
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,236
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-236-2

  Query Match             8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | |.|:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 9
US-09-221-416-2
; Sequence 2, Application US/09221416
; Patent No. 6153417
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,416
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-416-2

  Query Match             8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | |.|:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97
```

```
Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
              |||:||   : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 10
US-09-221-245-2
; Sequence 2, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,245
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: US 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-245-2

  Query Match           8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity  40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
              |||:||   : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 11
US-09-163-115-2
; Sequence 2, Application US/09163115A
; Patent No. 6183962
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/163,115A
;  CURRENT FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-163-115-2

  Query Match           8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity  40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
              |||:||   : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 12
US-09-221-528-2
; Sequence 2, Application US/09221528
; Patent No. 6190874
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,528
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-528-2

  Query Match           8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity  40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |      : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
```

```
                ||||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 13
US-09-593-553-2
; Sequence 2, Application US/09593553
; Patent No. 6200770
; GENERAL INFORMATION:
;  APPLICANT:  Acton, Susan
;  TITLE OF INVENTION:  NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE:  MNI-050
;  CURRENT APPLICATION NUMBER: US/09/593,553
;  CURRENT FILING DATE:  2000-06-14
;  PRIOR APPLICATION NUMBER:  09/163,115
;  PRIOR FILING DATE:  1998-09-28
;  NUMBER OF SEQ ID NOS:  15
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-593-553-2

  Query Match             8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |       : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 14
US-09-221-237-2
; Sequence 2, Application US/09221237
; Patent No. 6214597
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,237
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE: 1998-09-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-221-237-2

  Query Match             8.5%;  Score 441;  DB 4;  Length 302;
  Best Local Similarity   40.6%;  Pred. No. 2.6e-29;
  Matches  103;  Conservative   43;  Mismatches   98;  Indels   10;  Gaps    7;

Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |       : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


Qy         58 LGRGAFGEATLYRRT--EDDSLVVWKEVDLTRLSE-KERRDALNEIVILALLQHDNIIAY 114
              :||| | |  :||     |   |  |:| :  | : | | | : || :|  | | |:| |
Db         40 IGRGQFSE--VYRAACLLDGVPVALKKVQIFDLMDAKARADCIKEIDLLKQLNHPNVIKY 97

Qy        115 YNHFMDNTTLLIELEYCNGGNLYDKILR-QKDKLFEEEMVVW-YLFQIVSAVSCIHKAGI 172
              |  |:::  | | ||  : |:|   |   :| |    |  || |  |: ||:  :|   :
Db         98 YASFIEDNELNIVLELADAGDLSRMIKHFKKQKRLIPERTVWKYFVQLCSALEHMHSRRV 157

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              :|||||  |:|:|   ::|||| || :  :|: : | :||||||||||      |||||
Db        158 MHRDIKPANVFITATGVVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIHENGYNFKS 217

Qy        233 DIWAVGCVIFELLTLKRTF--DATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||::||:::|:  |:  |  |  |  :|| || |       : |  || || |:|: |::
Db        218 DIWSLGCLLYEMAALQSPFYGDKMNLYSLCKKIEQ-CDYPPLPSDHYSEELRQLVNMCIN 276

Qy        291 QDPEQRPTADELLD 304
               |||:||    : |
Db        277 PDPEKRPDVTYVYD 290


RESULT 15
US-08-712-709-1
; Sequence 1, Application US/08712709
; Patent No. 5863780
; GENERAL INFORMATION:
;   APPLICANT:  Au-Young, Janice
;   APPLICANT:  Guegler, Karl J.
;   APPLICANT:  Hawkins, Phillip R.
;   TITLE OF INVENTION:  NOVEL HUMAN PROTEIN KINASES
;   NUMBER OF SEQUENCES:  9
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;     STREET:  3174 Porter Drive
;     CITY:  Palo Alto
;     STATE:  CA
;     COUNTRY:  U.S.
;     ZIP:  94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/712,709
;     FILING DATE:  Filed Herewith
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Billings, Lucy J
;     REGISTRATION NUMBER:  36,749
;     REFERENCE/DOCKET NUMBER:  PF-0118 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415-855-0555
;     TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  233 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   IMMEDIATE SOURCE:
;     LIBRARY:
;     CLONE:  Consensus
US-08-712-709-1

  Query Match             7.8%;  Score 403.5;  DB 2;  Length 233;
  Best Local Similarity   40.2%;  Pred. No. 2.6e-26;
  Matches   86;  Conservative   40;  Mismatches   83;  Indels    5;  Gaps    3;

Qy         88 LSEKERRDALNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYD--KILRQKD 145
              :  | ::| : || :|  | | ||| | : |:::  | | ||  : |:|    |  :::
Db          2 MDAKAKQDCVKEIGLLKQLNHPNIIKYLDSFIEDNELNIVLELADAGDLPQMIKYFKKQK 61
```

```
Qy        146 KLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEY 205
              :|  |  |  |  |: |||  :|   ::|||||  |:|:|   ::|||| || :  :||
Db         62 RLIPERTVWKYFVQLCSAVEHMHSRRVMHRDIKPANVFITATGVVKLGDLGLGRFFSSET 121

Qy        206 SMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTF--DATNPLNLCVKI 263
              : | :|||||||||||      |||||||::||:::|:  |:  |  |  |  :|| ||
Db        122 TAAHSLVGTPYYMSPERIHENGYNFKSDIWSLGCLLYEMAALQSPFYGDKMNLFSLCQKI 181

Qy        264 VQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRP 297
               |      :    || :| ::|  |:  || |||
Db        182 EQ-CDYPPLPGEHYSEKLRELVSMCICPDPHQRP 214


Search completed: July  1, 2003, 08:36:00
Job time : 23.1764 secs
```

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:33:58 ; Search time 150.407 Seconds
(without alignments)
713.694 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters: 424699

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5183	100.0	979	9	US-09-884-001-4	Sequence 4, Appli
2	1084	20.9	692	10	US-09-992-481-2	Sequence 2, Appli
3	1081	20.9	692	10	US-09-910-150-5	Sequence 5, Appli
4	1062.5	20.5	713	9	US-10-162-706-2	Sequence 2, Appli
5	601	11.6	616	9	US-10-243-735-4	Sequence 4, Appli
6	601	11.6	774	9	US-10-162-706-5	Sequence 5, Appli
7	599.5	11.6	1214	10	US-09-783-320-4	Sequence 4, Appli
8	584	11.3	254	9	US-09-898-837A-35	Sequence 35, Appl
9	577	11.1	841	9	US-10-198-070-87	Sequence 87, Appl
10	572	11.0	841	9	US-10-198-070-52	Sequence 52, Appl
11	561	10.8	416	10	US-09-731-231A-6	Sequence 6, Appli
12	558	10.8	460	9	US-10-162-706-4	Sequence 4, Appli
13	553	10.7	280	9	US-10-162-706-6	Sequence 6, Appli
14	552.5	10.7	260	10	US-09-731-231A-5	Sequence 5, Appli
15	548.5	10.6	330	9	US-09-898-837A-9	Sequence 9, Appli
16	544.5	10.5	255	9	US-09-898-837A-36	Sequence 36, Appl
17	541	10.4	291	9	US-10-162-706-17	Sequence 17, Appl
18	538.5	10.4	640	9	US-10-243-735-2	Sequence 2, Appli
19	535.5	10.3	654	10	US-09-940-921B-4	Sequence 4, Appli
20	535.5	10.3	683	10	US-09-940-921B-2	Sequence 2, Appli
21	502.5	9.7	345	10	US-09-870-962-6	Sequence 6, Appli
22	499	9.6	445	9	US-10-242-943-2	Sequence 2, Appli
23	455.5	8.8	459	10	US-09-771-161A-198	Sequence 198, App
24	451	8.7	193	9	US-09-898-837A-40	Sequence 40, Appl
25	442.5	8.5	283	9	US-09-898-837A-34	Sequence 34, Appl
26	441	8.5	302	10	US-09-757-982-2	Sequence 2, Appli
27	441	8.5	311	9	US-10-106-698-6249	Sequence 6249, Ap
28	431.5	8.3	256	9	US-09-823-187-96	Sequence 96, Appl
29	429.5	8.3	645	10	US-09-731-231A-2	Sequence 2, Appli
30	410	7.9	4834	9	US-10-097-534-27	Sequence 27, Appl
31	407.5	7.9	249	9	US-09-764-868-812	Sequence 812, App
32	403.5	7.8	233	10	US-09-810-808-1	Sequence 1, Appli
33	403.5	7.8	467	9	US-09-866-050A-513	Sequence 513, App
34	386	7.4	892	10	US-09-887-828A-2	Sequence 2, Appli
35	368.5	7.1	239	10	US-09-810-808-7	Sequence 7, Appli
36	367	7.1	379	9	US-10-026-021-3	Sequence 3, Appli
37	367	7.1	970	9	US-10-026-021-2	Sequence 2, Appli
38	365	7.0	66	10	US-09-864-761-42067	Sequence 42067, A
39	357.5	6.9	912	9	US-09-291-417-26	Sequence 26, Appl
40	357.5	6.9	968	9	US-09-291-417-107	Sequence 107, App
41	357	6.9	4861	9	US-10-097-534-26	Sequence 26, Appl
42	357	6.9	4861	9	US-10-146-473-49	Sequence 49, Appl
43	357	6.9	4861	10	US-09-919-497-70	Sequence 70, Appl
44	356	6.9	367	9	US-10-026-021-6	Sequence 6, Appli
45	355.5	6.9	275	9	US-10-274-409-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-884-001-4
; Sequence 4, Application US/09884001
; Publication No. US20020182656A1
; GENERAL INFORMATION:
; APPLICANT: Bird, Timothy A.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Willis, Cynthia R.
; TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
; TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)
; FILE REFERENCE: Immunex GNK/sGNK PCT
; CURRENT APPLICATION NUMBER: US/09/884,001
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/113,003
; PRIOR FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-001-4

Query Match 100.0%; Score 5183; DB 9; Length 979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
```

```
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979

Db        961 DLDSDSWCLLGTDSCRPSL 979


RESULT 2
US-09-992-481-2
; Sequence 2, Application US/09992481
; Patent No. US20020110908A1
; GENERAL INFORMATION:
;  APPLICANT: Friddle, Carl Johan
;  APPLICANT:  Hilbun, Erin
;  APPLICANT:  Mathur, Brian
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020110908A1el Human Kinases and Polynucleotides
;  TITLE OF INVENTION:  Encoding the Same
;  FILE REFERENCE: LEX-0266-USA
;  CURRENT APPLICATION NUMBER: US/09/992,481
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: US 60/252,011
;  PRIOR FILING DATE: 2000-11-20
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 692
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-992-481-2

  Query Match            20.9%;  Score 1084;  DB 10;  Length 692;
  Best Local Similarity  36.7%;  Pred. No. 5.2e-69;
  Matches  260;  Conservative  121;  Mismatches  269;  Indels   58;  Gaps   17;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111

Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171

Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230

Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290

Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL--------------LRKRRREMEEKVTLLNAPTKRPRSSTVTE 336

Db        241 LEPAQRPPLSHIMAQPLCIRALLNLHTDVGSVRMRRAEKSVAPSNTGSRTTSVRCRGIPR 300

Qy        337 API-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW-- 393

Db        301 GPVRPAIPPPLSSVYAWGGGLGTPLRLPMLNT--EVVQVAAGRTQKAGVTRSGRLILWEA 358

Qy        394 --VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCVTDEGQL 447

Db        359 PPLGAGGGSLLPGAV------EQPQPQFISRFLEGQSGVTIKHVACGDFFTACLTDRGII 412

Qy        448 YAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCG 507

Db        413 MTFGSGSNGCLGHGSLT--DISQPTIVEALLGYEMVQVACGASHVLALSTERELFAWGRG 470

Qy        508 EYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLN 567

Db        471 DSGRLGLGTRESHSCPQQVPMPPGQEAQRVVCGIDSSMILTVPGQALACGSNRFNKLGLD 530

Qy        568 QCMSG--IINHEAYHEVPYTTSFTL--AKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGC 623

Db        531 HLSLGEEPVPHQ---QVEEALSFTLLGSAPLDQEPLLSIDLGTAHSAAVTASGDCYTFGS 587

Qy        624 NKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRL 682

Db        588 NQHGQLGTNTRRGSRAPCKVQG--LEGIKMAMVACGDAFTVAIGAESEVYSWGKGARGRL 645

Qy        683 AMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVEKV 727

Db        646 GR--------RDEDAGLPRPVQLDETHPYTVTSVSCCHGNTLLAVRSV 685


RESULT 3
US-09-910-150-5
; Sequence 5, Application US/09910150
; Patent No. US20020068698A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Meyers, Rachel
;  APPLICANT:  Rudolph-Owen, Laura
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Tsai, Fong Ying
;  TITLE OF INVENTION: 13237, 18480, 2245 OR 16228 NOVEL HUMAN
;  TITLE OF INVENTION:  PROTEIN KINASE MOLECULES AND USES THEREFOR
;  FILE REFERENCE: 38155-20020.00
```

```
;  CURRENT APPLICATION NUMBER: US/09/910,150
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/219,028
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 692
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-910-150-5

  Query Match            20.9%;  Score 1081;  DB 10;  Length 692;
  Best Local Similarity   36.7%;  Pred. No. 8.5e-69;
  Matches  260;  Conservative  120;  Mismatches  270;  Indels   58;  Gaps   17;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:||||||    |  |  |  ||: |:: : ::::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  | | :||  || | : | :: : | ||| :: :  :||: |:  :|
Db         64 IEYYENFLEDKALTIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  : :|:||:|::|  |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL-------------LRKRRREMEEKVTLLNAPTKRPRSSTVTE 336
               :| |||    :: :||              :| || |     :   : |   |   :
Db        241 LEPAQRPPLSHIMAQPLCIRALLNLHTDVGSVRMRRAEKSVAPSNTGSRTTSVRCRGIPR 300

Qy        337 API-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW-- 393
               |:   :    | || |||  || :| :: :     || || |  | ||    |  |
Db        301 GPVRPAIPPPLSSVYAWGGGLGTPLRLPMLNT--EVVQVAAGRTQKAGVTRSGRLILWEA 358

Qy        394 --VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCVTDEGQL 447
                :   ||: | | :          ||: : :      |  |  |:||| || |:|| | :
Db        359 PPLGAGGGSLLPGAV------EQPQPQFISRFLEGQSGVTIKHVACGDFFTACLTDRGII 412

Qy        448 YAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCG 507
               |||   ||:|   :   :: :|  :   |   : ||:|| :||: |:  :|:::|| |
Db        413 MTFGSGSNGCLGHGSLT--DISQPTIVEALLGYEMVQVACGASHVLALSTERELFAWGRG 470

Qy        508 EYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLN 567
              : ||||| : | :  ||:| :|       | || | : :||  |: |||| | ||||||:
Db        471 DSGRLGLGTRESHSCPQQVPMPPGQEAQRVVCGIDSSMILTVPGQALACGSNRFNKLGLD 530

Qy        568 QCMSG--IINHEAYHEVPYTTSFTL--AKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGC 623
                |   : |:   :|    |||| :  |    : :|  |  |:||:   |  |||
Db        531 HLSLGEEPVPHQ---QVEEALSFTLLGSAPLDQEPLLSIDLGTAHSAAVTASGDCYTFGS 587

Qy        624 NKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRL 682
              |: |||    :  |    : |  | | ::  |:||| ||:|   :: :::|| |  |||
Db        588 NQHGQLGTNTRRGSRAPCKVQG--LEGIKMAMVACGDAFTVAIGAESEVYSWGKGARGRL 645

Qy        683 AMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVEKV 727
                        |    |||:     |   |  :||   :|:| |  |
Db        646 GR--------RDEDAGLPRPVQLDETHPYTVTSVSCCHGNTLLAVRSV 685


RESULT 4
US-10-162-706-2
; Sequence 2, Application US/10162706
; Publication No. US20030059918A1
; GENERAL INFORMATION:
;  APPLICANT: Smolyar, Alex
;  TITLE OF INVENTION: Regulation of Human Serine/Threonine
;  TITLE OF INVENTION:  Kinase
;  FILE REFERENCE: 004974.00731
;  CURRENT APPLICATION NUMBER: US/10/162,706
;  CURRENT FILING DATE:  2002-06-06
;  PRIOR APPLICATION NUMBER: US 60/296,164
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/323,100
;  PRIOR FILING DATE: 2001-09-19
;  PRIOR APPLICATION NUMBER: US 60/330,578
;  PRIOR FILING DATE: 2001-10-25
;  PRIOR APPLICATION NUMBER: US 60/348,601
;  PRIOR FILING DATE: 2002-01-17
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 713
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-706-2

  Query Match            20.5%;  Score 1062.5;  DB 9;  Length 713;
  Best Local Similarity   35.8%;  Pred. No. 1.8e-67;
  Matches  261;  Conservative  124;  Mismatches  265;  Indels   79;  Gaps   20;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:|||||    |  |  |  ||: |:: : ::::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  : :|:||:|::|  |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL-LR------------KRRREMEEKVTLLNAPTKRPRSST---- 333
               :| |||    :: :|| :|            : || :: :  :|      |  ||
Db        241 LEPAQRPPLSHIMAQPLCIRALLNLHTDVGSVRMRRPVQGQRAVLGGRVWAPSGSTGGSR 300

Qy        334 --------VTEAPI-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVV 384
                     :   |:   :    | || ||||  || :| :: :      || || |  | |
Db        301 TTSVRCRGIPRGPVRPAIPPPLSSVYAWGGGLGTPLRLPMLNT--EVVQVAAGRTQKAGV 358

Qy        385 TVEKELYTW----VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDD 436
              |    |  |    :   ||: | | :          ||: : :      |  |  |:|||
Db        359 TRSGRLILWEAPPLGAGGGSLLPGAV------EQPQPQFISRFLEGQSGVTIKHVACGDF 412

Qy        437 FTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLT 496
              || |:|| | :  |||   ||:|   :   :: :|  :   |   : ||:|| :||: |:
Db        413 FTACLTDRGIIMTFGSGSNGCLGHGSLT--DISQPTIVEALLGYEMVQVACGASHVLALS 470

Qy        497 RNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLAC 556
               :|:::|| |: ||||| : | :  ||:| :|       | || | : :||  |: |||
Db        471 TERELFAWGRGDSGRLGLGTRESHSCPQQVPMPPGQEAQRVVCGIDSSMILTVPGQALAC 530

Qy        557 GLNE----------FNKLGLNQCMSG--IINHEAYHEVPYTTSFTL--AKQLSFYKIRTI 602
              | |           ||||||:    |   : |:   :|    |||| :  |    : :|
Db        531 GSNSWIWLLALPSGFNKLGLDHLSLGEEPVPHQ---QVEEALSFTLLGSAPLDQEPLLSI 587

Qy        603 APGKTHTAAIDERGRLLTFGCNKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEF 661
               |  |:||:   |   ||| |: ||||    :  |    : |  | | ::  |:||| |
Db        588 DLGTAHSAAVTASGDCYTFGSNQHGQLGTNTRRGSRAPCKVQG--LEGIKMAMVACGDAF 645

Qy        662 TIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGW 718
              |:|   :: :::|| |  |||          |    |||:     |   |  :||
Db        646 TVAIGAESEVYSWGKGARGRLGR--------RDEDAGLPRPVQLDETHPYTVTSVSCCHG 697

Qy        719 HTILIVEKV 727
              :|:| |  |
```

```
Db        698 NTLLAVRSV 706


RESULT 5
US-10-243-735-4
; Sequence 4, Application US/10243735
; Publication No. US20030022341A1
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001212DIV
;  CURRENT APPLICATION NUMBER: US/10/243,735
;  CURRENT FILING DATE:  2002-09-16
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 616
;   TYPE: PRT
;   ORGANISM: Mouse
US-10-243-735-4

  Query Match             11.6%;  Score 601;  DB 9;  Length 616;
  Best Local Similarity   39.8%;  Pred. No. 1.3e-34;
  Matches  117;  Conservative   67;  Mismatches   94;  Indels   16;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::|:|:|||::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYKESFEENGSLYIVMDYCEGGDLFKRINAQKGALFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| |  || :||: |  :    |  || :|  ::    :
Db        184 SDIWALGCVLYELCTLKHAFEAGNMKNLVLKIISG--SFPPVSPHYSYDLRSLLSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRRE------MEEKVTLLN--------APTKRPRS 331
              :|  ||: : :|::  : ||  :      : |:  |           | ||| |
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTLSKFGPQPLPGKRPAS 295


RESULT 6
US-10-162-706-5
; Sequence 5, Application US/10162706
; Publication No. US20030059918A1
; GENERAL INFORMATION:
;  APPLICANT: Smolyar, Alex
;  TITLE OF INVENTION: Regulation of Human Serine/Threonine
;  TITLE OF INVENTION:  Kinase
;  FILE REFERENCE: 004974.00731
;  CURRENT APPLICATION NUMBER: US/10/162,706
;  CURRENT FILING DATE:  2002-06-06
;  PRIOR APPLICATION NUMBER: US 60/296,164
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/323,100
;  PRIOR FILING DATE: 2001-09-19
;  PRIOR APPLICATION NUMBER: US 60/330,578
;  PRIOR FILING DATE: 2001-10-25
;  PRIOR APPLICATION NUMBER: US 60/348,601
;  PRIOR FILING DATE: 2002-01-17
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 774
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-706-5

  Query Match             11.6%;  Score 601;  DB 9;  Length 774;
  Best Local Similarity   39.8%;  Pred. No. 1.8e-34;
  Matches  117;  Conservative   67;  Mismatches   94;  Indels   16;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::|:|:|||::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYKESFEENGSLYIVMDYCEGGDLFKRINAQKGALFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| |  || :||: |  :    |  || :|  ::    :
Db        184 SDIWALGCVLYELCTLKHAFEAGNMKNLVLKIISG--SFPPVSPHYSYDLRSLLSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRRE------MEEKVTLLN--------APTKRPRS 331
              :|  ||: : :|::  : ||  :      : |:  |           | ||| |
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTLSKFGPQPLPGKRPAS 295


RESULT 7
US-09-783-320-4
; Sequence 4, Application US/09783320
; Patent No. US20020038011A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Hu, Yi
;  APPLICANT:  Nepomnichy, Boris
;  APPLICANT:  Turner, C. Alexander Jr
;  APPLICANT:  Zambrowicz, Brian
;  TITLE OF INVENTION: No. US20020038011Alel Human Kinases and Polynucleotides Encodi
;  FILE REFERENCE: LEX-0137-USA
;  CURRENT APPLICATION NUMBER: US/09/783,320
;  CURRENT FILING DATE:  2001-02-15
;  PRIOR APPLICATION NUMBER: US 60/183,582
;  PRIOR FILING DATE: 2000-02-18
;  PRIOR APPLICATION NUMBER: US 60/184,014
;  PRIOR FILING DATE: 2000-02-22
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1214
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-783-320-4

  Query Match             11.6%;  Score 599.5;  DB 10;  Length 1214;
  Best Local Similarity   37.5%;  Pred. No. 4.4e-34;
  Matches  124;  Conservative   71;  Mismatches  113;  Indels   23;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| |  || :||: |  :    |  || :|  ::    :
Db        184 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 241
```

```
Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :       ||       |     | ||| |  :::: 
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 301

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::      |   | |   :| 
Db        302 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 332


RESULT 8
US-09-898-837A-35
; Sequence 35, Application US/09898837A
; Publication No. US20030077697A1
; GENERAL INFORMATION:
;  APPLICANT: Quinn, Kerry E.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Herrmann, John L.
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Taupier Jr., Raymond
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  CuraGen Corporation
;  APPLICANT:  Gerlach, Valerie L
;  APPLICANT:  MacDougall, John R
;  TITLE OF INVENTION: NOVEL SERINE/THREONINE PROTEIN-KINASE LIKE PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING THE SAME
;  FILE REFERENCE: 15966-598 CIP
;  CURRENT APPLICATION NUMBER: US/09/898,837A
;  CURRENT FILING DATE:  2001-07-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
;  PRIOR FILING DATE: 1999-11-17
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
;  PRIOR FILING DATE: 2000-04-05
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
;  PRIOR FILING DATE: 2000-02-09
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
;  PRIOR FILING DATE: 2000-04-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/215,906
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 09/715,427
;  PRIOR FILING DATE: 2000-11-16
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 254
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-898-837A-35

  Query Match           11.3%;  Score 584;  DB 9;  Length 254;
  Best Local Similarity  42.3%;  Pred. No. 6.4e-34;
  Matches  107;  Conservative   63;  Mismatches   81;  Indels    2;  Gaps    1;

Qy         53 IPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNII 112
              : :: :| |:||:| | : |||    | ||:::::|:|:|||::  |: :|| ::| ||: 
Db          1 VRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNIV 60

Qy        113 AYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGI 172
                |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|   |
Db         61 QYKESFEENGSLYIVMDYCEGGDLFKRINAQKGALFQEDQILDWFVQICLALKHVHDRKI 120

Qy        173 LHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKS 232
              ||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   |.| ||
Db        121 LHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNKS 180

Qy        233 DIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQD 292
              ||||:|||::|| |||  |:| |  || :|||: |  :    |  || :|  ::     ::
Db        181 DIWALGCVLYELCTLKHAFEAGNMKNLVLKIISG--SFPPVSPHYSYDLRSLLSQLFKRN 238

Qy        293 PEQRPTADELLDR 305
              |  ||: : :|::
Db        239 PRDRPSVNSILEK 251


RESULT 9
US-10-198-070-87
; Sequence 87, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT:  GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
;   LENGTH: 841
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-198-070-87

  Query Match           11.1%;  Score 577;  DB 9;  Length 841;
  Best Local Similarity  35.4%;  Pred. No. 1.1e-32;
  Matches  128;  Conservative   71;  Mismatches  131;  Indels   32;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|:|:::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYLRVVGKGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|: 
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPENQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:||||: |:||:|| |:|: | :   || ||:||||||||||     ||: 
Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHCDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDSSQYSLELIQMVHSCL 289
              |||:||:|| ::|: |||  |:| :  :|  :|::| : ||  |   || ||  ::: : |
Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEGKLPAMPRD---YSPELAELIRTML 242

Qy        290 DQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVV------- 342
              : ||:||:   :| :| : ||:     : | :       ::     . | | |
Db        243 SKRPEERPSVRSILRQPYI-KRQISFFLEATKIKTSKNNIKNGDSQSKPFATVVSGEAES 301

Qy        343 ----------TSRTSEVYVWGGGK----STPQKLDVIKSGCS-----ARQVCAGNTHFAV 383
                       :|  |: |: | ||      |:   ::||  |      :|  :  |  |
Db        302 NHEVIHPQPLSSEGSQTYIMGEGKCLSQEKPRASGLLKSPASLKAHTCKQDLSNTTELAT 361

Qy        384 VT 385
              ::
Db        362 IS 363


RESULT 10
US-10-198-070-52
; Sequence 52, Application US/10198070
; Publication No. US20030109437A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT: GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
;   LENGTH: 841
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-198-070-52

  Query Match            11.0%;  Score 572;  DB 9;  Length 841;
  Best Local Similarity   35.1%;  Pred. No. 2.4e-32;
  Matches  127;  Conservative   71;  Mismatches  132;  Indels   32;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|:|:::|| || :    |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYLRVVGKGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :      || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPENQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:|||: |:||:|| |:|: | :    || ||:|||||||||     ||:
Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHCDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDSSQYSLELIQMVHSCL 289
              |||:||:|| ::|: |||  |:| :  :|  :|:::| :  |  |   || || ::: : |
Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEGKLPPMPRD---YSPELAELIRTML 242

Qy        290 DQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVV------- 342
              : ||:||:   :| :| : ||:     : | :      ::       | | |
Db        243 SKRPEERPSVRSILRQPYI-KRQISFFLEATKIKTSKNNIKNGDSQSKPFATVVSGEAES 301

Qy        343 ----------TSRTSEVYVWGGGK----STPQKLDVIKSGCS-----ARQVCAGNTHFAV 383
                       :|  |: |: | ||      |:   ::||  |      :|  :  |  |
Db        302 NHEVIHPQPLSSEGSQTYIMGEGKCLSQEKPRASGLLKSPASLKAHTCKQDLSNTTELAT 361

Qy        384 VT 385
              ::
Db        362 IS 363


RESULT 11
US-09-731-231A-6
; Sequence 6, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007
;  CURRENT APPLICATION NUMBER: US/09/731,231A
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-731-231A-6

  Query Match            10.8%;  Score 561;  DB 10;  Length 416;
  Best Local Similarity   33.0%;  Pred. No. 5.6e-32;
  Matches  140;  Conservative   72;  Mismatches  146;  Indels   66;  Gaps   11;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|||::|| || :    |    | |:::|   | :||| |  |  :|: |:| ||
Db          1 YCYMRVVGRGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 60

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :      || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         61 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPESQVVEWFVQIAMALQYLHEK 120

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:|||: |:||:|| |:|: | :    || ||:|||||||||     ||:
Db        121 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHGDMASTLIGTPYYMSPELFSNKPYNY 180

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||:||:|| ::|: |||  |:| :  :|  :|:::|   :       || || ::: : |
Db        181 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEG--KLPPMPKVYSTELAELIRTMLS 238

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTE------APIAVVTS 344
              : ||:||:   :| :| ::          ::|    ||   |    :       |:| | |
Db        239 RRPEERPSVRSILRQPYIK-------HHISLFLEATKAKTSKNNVKNCDSRAKPVAAVVS 291

Qy        345 RTSE----------------VYVWGGGKSTPQKLDV----IKSGCSAR----QVCAGNT 379
              |  |                  ::| |  |   |:  |    ::|  |      :    ||
Db        292 RKEESNTDVIHYQPRSSEGSALHVMGEDKCLSQEKPVDIGPLRSPASLEGHTGKQDMNNT 351

Qy        380 HFAVVTVEKELYTWVNM-----------QGGTKLHGQLGHGDKA-------SYRQPKHVE 421
                :  |: :     :|:           |     |  | | |       |  | | |
Db        352 GESCATISR-----INIDILPAERRDSANAGVVQESQPQHVDAADEVDSQCSISQEK--E 404

Qy        422 KLQG 425
              :|||
Db        405 RLQG 408


RESULT 12
US-10-162-706-4
; Sequence 4, Application US/10162706
; Publication No. US20030059918A1
; GENERAL INFORMATION:
;  APPLICANT: Smolyar, Alex
;  TITLE OF INVENTION: Regulation of Human Serine/Threonine
;  TITLE OF INVENTION:  Kinase
;  FILE REFERENCE: 004974.00731
;  CURRENT APPLICATION NUMBER: US/10/162,706
;  CURRENT FILING DATE:  2002-06-06
;  PRIOR APPLICATION NUMBER: US 60/296,164
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/323,100
;  PRIOR FILING DATE: 2001-09-19
;  PRIOR APPLICATION NUMBER: US 60/330,578
;  PRIOR FILING DATE: 2001-10-25
;  PRIOR APPLICATION NUMBER: US 60/348,601
;  PRIOR FILING DATE: 2002-01-17
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-706-4

  Query Match            10.8%;  Score 558;  DB 9;  Length 460;
  Best Local Similarity   41.3%;  Pred. No. 1e-31;
  Matches  130;  Conservative   58;  Mismatches   97;  Indels   30;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
                 |  |||:||||||   |  |  |  ||: |:: : :::::||: | ||  :| || | |:
```

```
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              ||||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPT-------KRPRSSTVTEAPIAVVT 343
               :| |||    :: :||  :          |||  |       :||      :   ||:
Db        241 LEPAQRPPLSHIMAQPLCIR---------ALLNLHTDVGSVRMRRP-----VQGQRAVLG 286

Qy        344 SRTSEVYVWGGGKST 358
              |      ||    ||
Db        287 GR-----VWAPSGST 296


RESULT 13
US-10-162-706-6
; Sequence 6, Application US/10162706
; Publication No. US20030059918A1
; GENERAL INFORMATION:
;  APPLICANT: Smolyar, Alex
;  TITLE OF INVENTION: Regulation of Human Serine/Threonine
;  TITLE OF INVENTION:  Kinase
;  FILE REFERENCE: 004974.00731
;  CURRENT APPLICATION NUMBER: US/10/162,706
;  CURRENT FILING DATE:  2002-06-06
;  PRIOR APPLICATION NUMBER: US 60/296,164
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/323,100
;  PRIOR FILING DATE: 2001-09-19
;  PRIOR APPLICATION NUMBER: US 60/330,578
;  PRIOR FILING DATE: 2001-10-25
;  PRIOR APPLICATION NUMBER: US 60/348,601
;  PRIOR FILING DATE: 2002-01-17
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 280
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-706-6

  Query Match             10.7%;  Score 553;  DB 9;  Length 280;
  Best Local Similarity   45.5%;  Pred. No. 1.2e-31;
  Matches  117;  Conservative   54;  Mismatches   82;  Indels    4;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:||||||    |  |  |  ||: |:: : :::::||: | || :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL 307
               :| |||    :: :||
Db        241 LEPAQRPPLSHIMAQPL 257


RESULT 14
US-09-731-231A-5
; Sequence 5, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007
;  CURRENT APPLICATION NUMBER: US/09/731,231A
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Human
US-09-731-231A-5

  Query Match             10.7%;  Score 552.5;  DB 10;  Length 260;
  Best Local Similarity   41.6%;  Pred. No. 1.2e-31;
  Matches  109;  Conservative   59;  Mismatches   89;  Indels    5;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|:|::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          1 YCYLRVVGKGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 60

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         61 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPENQVVEWFVQIAMALQYLHEK 120

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| |:|||: |:||:|| |:|: | :   || ||:||||||||||     ||:
Db        121 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHCDMASTLIGTPYYMSPELFSNKPYNY 180

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDSSQYSLELIQMVHSCL 289
              |||:||:|| ::|: |||  |:| :  :|  :|::| : ||  |   || || ::: : |
Db        181 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEGKLPAMPRD---YSPELAELIRTML 237

Qy        290 DQDPEQRPTADELLDRPLLRKR 311
               : ||:||:   :| :| ::::
Db        238 SKRPEERPSVRSILRQPYIKRQ 259


RESULT 15
US-09-898-837A-9
; Sequence 9, Application US/09898837A
; Publication No. US20030077697A1
; GENERAL INFORMATION:
;  APPLICANT: Quinn, Kerry E.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Herrmann, John L.
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Taupier Jr., Raymond
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  CuraGen Corporation
;  APPLICANT:  Gerlach, Valerie L
;  APPLICANT:  MacDougall, John R
;  TITLE OF INVENTION: NOVEL SERINE/THREONINE PROTEIN-KINASE LIKE PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING THE SAME
;  FILE REFERENCE: 15966-598 CIP
;  CURRENT APPLICATION NUMBER: US/09/898,837A
;  CURRENT FILING DATE:  2001-07-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
;  PRIOR FILING DATE: 1999-11-17
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
;  PRIOR FILING DATE: 2000-04-05
```

```
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
;  PRIOR FILING DATE: 2000-02-09
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
;  PRIOR FILING DATE: 2000-04-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 60/215,906
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: U.S.S.N. 09/715,427
;  PRIOR FILING DATE: 2000-11-16
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 330
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-898-837A-9

  Query Match             10.6%;  Score 548.5;  DB 9;  Length 330;
  Best Local Similarity   37.4%;  Pred. No. 3.1e-31;
  Matches  135;  Conservative   66;  Mismatches  113;  Indels   47;  Gaps   11;

Qy         46 EQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILAL 105
              | ||     :|| |||||   |  |  |  ||: |:: : :::::||: | ||  :| |
Db          2 ETEE---QVRVKGRGAFGIVQLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKL 57

Qy        106 LQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVS 165
              | | |:| || :|::: . |:| :||  || | : | :: : | ||| :: :  ||: |:
Db         58 LNHPNVIEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALH 117

Qy        166 CIHKAGILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQ 224
               :|   |||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:||||:
Db        118 HVHTHLILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCE 176

Qy        225 GVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVD------SSQYS 278
              |  || ||||||:|||::|| :||| |:|   :::|   :||             | :||
Db        177 GKPYNQKSDIWALGCVLYELASLKRAFEAA--VSVCT--LQGTTEKSTASPSSPLSDRYS 232

Qy        279 LELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPT-------KRPRS 331
               || |:| | |  :| |||    :: :||  :          |||  |       :||
Db        233 PELRQLVLSLLSLEPAQRPPLSHIMAQPLCIR---------ALLNLHTDVGSVRMRRP-- 281

Qy        332 STVTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELY 391
              :   ||:  |     ||    ||   | |  : |:        : | : |:  :|
Db        282 ---VQGQRAVLGGR-----VWAPSGSTLSPLTVSATACTYTL-----SSFTIDTLHHDLK 328

Qy        392 T 392
              |
Db        329 T 329


Search completed: July  1, 2003, 08:58:09
Job time : 152.407 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 43.4387 Seconds
(without alignments)
4643.787 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5183	100.0	979	4	Q8TD19	Q8td19 homo sapien
2	5169	99.7	979	4	Q8TCY4	Q8tcy4 homo sapien
3	2223	42.9	414	4	Q9Y6S5	Q9y6s5 homo sapien
4	1627	31.4	312	4	Q9Y6S4	Q9y6s4 homo sapien
5	1073	20.7	698	11	Q91ZR4	Q91zr4 mus musculu
6	1052.5	20.3	697	13	Q90XC2	Q90xc2 brachydanio
7	1034	19.9	198	4	Q9Y6S6	Q9y6s6 homo sapien
8	865	16.7	196	11	Q8R3P1	Q8r3p1 mus musculu
9	848	16.4	841	5	Q9VC32	Q9vc32 drosophila
10	670.5	12.9	579	5	Q95XQ3	Q95xq3 caenorhabdi
11	599.5	11.6	1265	4	Q96PY6	Q96py6 homo sapien
12	561	10.8	744	11	O35673	O35673 mus musculu
13	561	10.8	792	11	Q9R1J1	Q9r1j1 mus musculu
14	561	10.8	792	11	Q9Z1J2	Q9z1j2 mus musculu
15	541	10.4	291	11	Q9D685	Q9d685 mus musculu
16	524	10.1	555	5	Q9N9C3	Q9n9c3 leishmania
17	503.5	9.7	443	11	Q91Z18	Q91z18 mus musculu
18	499	9.6	384	4	Q96QN9	Q96qn9 homo sapien
19	497.5	9.6	509	11	Q99K72	Q99k72 mus musculu
20	495	9.6	442	13	Q9W622	Q9w622 xenopus lae
21	494.5	9.5	443	11	Q921N9	Q921n9 mus musculu
22	494	9.5	389	13	Q9W623	Q9w623 xenopus lae
23	475	9.2	393	11	Q91XQ1	Q91xq1 rattus norv
24	452.5	8.7	609	10	Q947T1	Q947t1 lycopersico
25	448	8.6	568	10	Q8RX66	Q8rx66 arabidopsis
26	447.5	8.6	735	5	Q9W3N8	Q9w3n8 drosophila
27	441	8.5	302	4	Q8TDX7	Q8tdx7 homo sapien
28	439	8.5	302	11	Q9ES74	Q9es74 mus musculu
29	439	8.5	555	10	Q8RXT4	Q8rxt4 arabidopsis
30	439	8.5	722	3	O13839	O13839 schizosacch
31	432.5	8.3	606	10	Q9CAU7	Q9cau7 arabidopsis
32	432	8.3	306	4	Q9ULX2	Q9ulx2 homo sapien
33	432	8.3	313	4	Q96E83	Q96e83 homo sapien
34	432	8.3	338	4	Q9HC98	Q9hc98 homo sapien
35	431	8.3	306	11	Q9DBI8	Q9dbi8 mus musculu
36	430	8.3	299	5	Q8T755	Q8t755 branchiosto
37	430	8.3	306	11	Q9D0E2	Q9d0e2 mus musculu
38	430	8.3	313	11	Q9ES70	Q9es70 mus musculu
39	429.5	8.3	482	4	Q8TBY1	Q8tby1 homo sapien
40	428.5	8.3	618	10	Q8S3U7	Q8s3u7 chlamydomon
41	421.5	8.1	637	6	Q8WNU8	Q8wnu8 macaca fasc
42	420	8.1	416	10	Q9LT35	Q91t35 arabidopsis
43	419	8.1	943	10	Q94CU5	Q94cu5 oryza sativ
44	417.5	8.1	621	10	Q8SA64	Q8sa64 populus x c
45	417	8.0	440	10	Q9XHD7	Q9xhd7 arabidopsis

ALIGNMENTS

RESULT 1
Q8TD19
ID Q8TD19 PRELIMINARY; PRT; 979 AA.
AC Q8TD19;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE NIMA-related kinase Nek8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland P.M., Milne A., Garka K., Johnson R.S., Willis C.R.,
RA Sims J.E., Rauch C.T., Bird T.A., Virca G.D.;
RT "Purification, cloning and characterization of a novel NIMA-related
RT kinase, Nek8, and its candidate substrate Bicd2.";
RL J. Biol. Chem. 0:0-0(2002).
DR EMBL; AY048580; AAL05428.1; -.
KW Kinase.
SQ SEQUENCE 979 AA; 107149 MW; FF2486CC599322CE CRC64;

Query Match 100.0%; Score 5183; DB 4; Length 979;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 979; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
```

```
Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979
              |||||||||||||||||||
Db        961 DLDSDSWCLLGTDSCRPSL 979
```

```
RESULT 2
Q8TCY4
ID   Q8TCY4      PRELIMINARY;      PRT;   979 AA.
AC   Q8TCY4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NIMA-family kinase NERCC.
GN   NERCC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Roig J., Mikhailov A., Belham C., Avruch J.;
RT   "Nercc, a mammalian NIMA-family kinase, binds the Ran GTPase and
RT   regulates mitotic progression.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY080896; AAL87410.1; -.
KW   Kinase.
SQ   SEQUENCE   979 AA;  107034 MW;  002483C1711DBADA CRC64;

  Query Match           99.7%;  Score 5169;  DB 4;  Length 979;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches  977;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        301 ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYIWGGGKSTPQ 360

Qy        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420

Qy        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480

Qy        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540

Qy        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600

Qy        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660

Qy        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720

Qy        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ILIVEKVLNSKTIRSNSSGLSIGTVFQSSSPGGGGGGGGGEEEDSQQESETPDPSGGFRG 780

Qy        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TMEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEK 840

Qy        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 DTLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVE 900

Qy        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        901 RLQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKP 960

Qy        961 DLDSDSWCLLGTDSCRPSL 979
              ||||||  ||||||||||||
Db        961 DLDSDSGCLLGTDSCRPSL 979


RESULT 3
Q9Y6S5
ID   Q9Y6S5      PRELIMINARY;      PRT;   414 AA.
AC   Q9Y6S5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similarity is to the end of 'Herc2.' (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dickhoff R., James R.,
RA   Loretz C., Lasky S., Madan A., Prescott S., Ratcliffe A., Shaffer T.,
RA   Hood L.;
RT   "Sequencing of human chromosome 14q24.3 region.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007055; AAD31939.1; -.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   Pfam; PF00415; RCC1; 2.
DR   PRINTS; PR00633; RCCNDNSATION.
DR   PROSITE; PS50012; RCC1_3; 6.
FT   NON_TER       1       1
FT   NON_TER     414     414
SQ   SEQUENCE   414 AA;  44876 MW;  9F938B4C2ED8871A CRC64;

  Query Match             42.9%;  Score 2223;  DB 4;  Length 414;
  Best Local Similarity   99.8%;  Pred. No. 4.5e-160;
  Matches  413;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        331 SSTVTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKEL 390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SSTVTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKEL 60

Qy        391 YTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAF 450
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 YTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIRQVSCGDDFTVCVTDEGQLYAF 120

Qy        451 GSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCGEYG 510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCGEYG 180

Qy        511 RLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQCM 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQCM 240

Qy        571 SGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQLG 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQLG 300

Qy        631 VGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTERP 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTERP 360

Qy        691 HGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLSIGT 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNSKTIRSNSSGLSIGT 414


RESULT 4
Q9Y6S4
ID   Q9Y6S4      PRELIMINARY;      PRT;   312 AA.
AC   Q9Y6S4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Similarity is to serine/threonine-protein kinase (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dickhoff R., James R.,
RA   Loretz C., Lasky S., Madan A., Prescott S., Ratcliffe A., Shaffer T.,
RA   Hood L.;
RT   "Sequencing of human chromosome 14q24.3 region.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AC007055; AAD31940.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Kinase; Serine/threonine-protein kinase.
FT   NON_TER     312     312
SQ   SEQUENCE   312 AA;  35283 MW;  A7C5DAA1E9E3F9A9 CRC64;

  Query Match             31.4%;  Score 1627;  DB 4;  Length 312;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-115;
  Matches  312;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR 60

Qy         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD 120

Qy        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL 180

Qy        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV 240

Qy        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD 300

Qy        301 ELLDRPLLRKRR 312
              ||||||||||||
Db        301 ELLDRPLLRKRR 312


RESULT 5
Q91ZR4
ID   Q91ZR4      PRELIMINARY;      PRT;   698 AA.
AC   Q91ZR4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NIMA-related kinase 8.
GN   NEK8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=C57BL/6J;
RA   Liu S., Lu W., Obara-Ishihara T., Drummond I., Beier D.R.;
RT   "A defect in a novel Nek-family kinase causes cystic disease in the
RT   mouse and in zebrafish.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF407579; AAL09675.1; -.
DR   MGD; MGI:1890646; Nek8.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00415; RCC1; 3.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR   PROSITE; PS50012; RCC1_3; 5.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   698 AA;  75264 MW;  686B29A8CF180E94 CRC64;

  Query Match             20.7%;  Score 1073;  DB 11;  Length 698;
  Best Local Similarity   36.8%;  Pred. No. 1.6e-72;
  Matches  263;  Conservative  121;  Mismatches  266;  Indels   64;  Gaps   19;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:||||||    |  | :|   ||: |:: : :::::||: | ||  :| || | |:
Db          4 YERIRVVGRGAFGIVHLCLRKADQKLVILKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  : :|:||:|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPLLRK-----------RRREMEEKVT----LLNAPTKRPRSST- 333
               :| |||    :: :|| :            | |  |: :|    : :  |    :|
Db        241 LEPAQRPPLSHIMAQPLCIRALLNIHTDVGSVRMRRAEKSLTPGPPIASGSTGSRATSAR 300

Qy        334 ---VTEAPI-AVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKE 389
                 |   |:   :    | || |||| |:| :| :: :     || || |  | ||
Db        301 CRGVPRGPVRPAIPPPLSSVYAWGGGLSSPLRLPMLNT--EVVQVAAGRTQKAGVTRSGR 358

Qy        390 LYTW----VNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCV 441
              | |    :  ||| | | :          ||: | :      |  |  |:||| || |:
Db        359 LILWEAPPLGAGGGTLLPGAV------ELPQPQFVSRFLEGQSGVTIKHVACGDLFTACL 412

Qy        442 TDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEV 501
              || | :  |||   ||:|   :   :: :|  :   |   : ||:|| :||: |: : |:
Db        413 TDRGIIMTFGSGSNGCLGHGNLT--DISQPTIVEALLGYEMVQVACGASHVLALSTDGEL 470

Qy        502 YSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEF 561
              ::|| |: ||||| : | :  ||:| |         | || | : :|| |:|||||  | |
Db        471 FAWGRGDGGRLGLGTRESHNCPQQVPVAPGQEAQRVVCGIDSSMILTSPGRVLACGSNRF 530

Qy        562 NKLGLNQCMSGIINHE--AYHEVPYTTSFTL--AKQLSFYKIRTIAPGKTHTAAIDERGR 617
              |||||:      :: |   | :|    |||   :  |     :  :  |  |:|||   |
Db        531 NKLGLDHLS---LDEEPVPYQQVEEALSFTPLGSAPLDQEPLLCVDLGTAHSAAITASGD 587

Qy        618 LLTFGCNKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGN 676
              |||  |: ||||  : :  |    : |  | | ::: |:||| ||:|   :  :::||
Db        588 CYTFGSNQHGQLGTSSRRVSRAPCRVQG--LEGIKMVMVACGDAFTVAVGAEGEVYSWGK 645

Qy        677 GGNGRLAMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVEKV 727
              |  |||           |     |||:     |   |  :||   :|:| |  |
Db        646 GTRGRLGR--------RDEDAGLPRPVQLDETHPYMVTSVSCCHGNTLLAVRSV 691
```

```
RESULT 6
Q90XC2
ID   Q90XC2      PRELIMINARY;      PRT;   697 AA.
AC   Q90XC2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NIMA-related kinase 8.
GN   NEK8.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu S., Lu W., Obara-Ishihara T., Drummond I., Beier D.R.;
RT   "A defect in a novel Nek-family kinase causes cystic disease in the
RT   mouse and in zebrafish.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF407580; AAL09676.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00415; RCC1; 3.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR   PROSITE; PS50012; RCC1_3; 4.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   697 AA;  76541 MW;  875A51D1F3831AA7 CRC64;

  Query Match             20.3%;  Score 1052.5;  DB 13;  Length 697;
  Best Local Similarity   36.7%;  Pred. No. 5.6e-71;
  Matches  261;  Conservative  128;  Mismatches  259;  Indels   63;  Gaps   21;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
                |   :|:|||||   | ||  | :||: ||: : :::  ||  | ||  :| || | ||
Db          4 YEKTKVVGRGAFGIVHLCRRRTDSALVILKEIPVEQMTRDERLAAQNECQVLKLLSHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | | | :: : | :|: ::    ||: |:  :| .
Db         64 IEYYENFLEDKALMIAMEYAPGGTLADYIQKRCNSLLDEDTILHSFVQILLALYHVHNKL 123

Qy        172 ILHRDIKTLNIFLTKANLI-KLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| || | |  :| |:||:|:::| | |: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHQMIVKIGDFGISKILVSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:: : |:
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLILNMLN 240

Qy        291 QDPEQRPTADELLD-----RPLL-------RKRRREMEEKVTLLNA-PTKRPR---SSTV 334
               || :||  :|::      ||||          : | :|: :: : | |  ||    :||
Db        241 LDPSKRPQLNEIMAHAICIRPLLNLYTDIGNVKMRRIEKPLSNVQAGPHGRPGGWITSTR 300

Qy        335 TEAPIAVVTSRTS-------EVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVE 387
              |   :: :||           || || | ||| :| :: :     ||  | |    ||
Db        301 TRGGLSSLTSSKMMHPLPLFSVYTWGSGISTPLRLPMLNT--EVIQVSLGRTQKMGVTKS 358

Qy        388 KELYTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKL----QGKAIHQVSCGDDFTVCVTD 443
              | ||     |:   |:          ||: : :      |  |  ||||| || |:||
Db        359 GRLITWEAPSVGS---GEPTLPGAVEQMQPQFISRFLEGQSGVTIKSVSCGDLFTTCLTD 415

Qy        444 EGQLYAFGSDYYGCMG---VDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKE 500
              | :  |||   ||:|     : |  |:::| :      |   : ||||| :||: :|  :|
Db        416 RGIIMTFGSGSNGCLGHGNFNDVTQPKIVEAL-----LGYELVQVSCGASHVLAVTNERE 470

Qy        501 VYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNE 560
              |:||| |: ||||| ::: :  ||:| :|       | || | : :::   ::|||| |
Db        471 VFSWGRGDNGRLGLATQDSHNCPQQVSLPADFEAQRVLCGVDCSMIMSTQHQILACGNNR 530
```

```
Qy        561 FNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQ--LSFYKIRTIAPGKTHTAAIDERGRL 618
              ||||||:: :||     :: :|     | | :   |:  ||  |  |  |: |: |:|:
Db        531 FNKLGLDK-VSGTEEPSSFCQVEEVHLFQLVQSAPLNTEKIVYIDIGTAHSVAVTEKGQC 589

Qy        619 LTFGCNKCGQLGVGNYK-KRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNG 677
              ||| |: |||| : :  |:   : |  | |  :   :||| ||:|   :  :: || |
Db        590 FTFGSNQHGQLGCSHRRSSRVPYQVSG--LQG--ITMAACGDAFTLAIGAEGEVYTWGKG 645

Qy        678 GNGRLAMTPTERPHGSDICTSWPRPIFGSLHH---VPDLSCRGWHTILIVE 725
                |||     |   |        |:|:     |   |  ::|   :|:| |:
Db        646 ARGRLGR--KEEDFGI------PKPVQLDESHAFTVTSVACCHGNTLLAVK 688


RESULT 7
Q9Y6S6
ID   Q9Y6S6      PRELIMINARY;      PRT;   198 AA.
AC   Q9Y6S6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Unknown (protein for MGC:16714).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dickhoff R., James R.,
RA   Loretz C., Lasky S., Madan A., Prescott S., Ratcliffe A., Shaffer T.,
RA   Hood L.;
RT   "Sequencing of human chromosome 14q24.3 region.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007055; AAD31938.1; -.
DR   EMBL; BC009336; AAH09336.1; -.
SQ   SEQUENCE   198 AA;  21384 MW;  8CE655EE50D6A92D CRC64;

  Query Match             19.9%;  Score 1034;  DB 4;  Length 198;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-70;
  Matches  198;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        782 MEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEKD 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEKD 60

Qy        842 TLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVER 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVER 120

Qy        902 LQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKPD 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKPD 180

Qy        962 LDSDSWCLLGTDSCRPSL 979
              ||||||||||||||||||
Db        181 LDSDSWCLLGTDSCRPSL 198


RESULT 8
Q8R3P1
ID   Q8R3P1      PRELIMINARY;      PRT;   196 AA.
AC   Q8R3P1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to hypothetical protein MGC16714.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024926; AAH24926.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   196 AA;  21214 MW;  E90468077DC330C4 CRC64;

  Query Match             16.7%;  Score 865;  DB 11;  Length 196;
  Best Local Similarity   85.4%;  Pred. No. 1.3e-57;
  Matches  169;  Conservative   10;  Mismatches   17;  Indels    2;  Gaps    1;

Qy        782 MEADRGMEGLISPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDSPSPLSAAFSESEKD 841
              |||||||||||||||||:||| |||||||||||||||||||||||||:|||||||:||||
Db          1 MEADRGMEGLISPTEAVGNSCGASSSCPGWLRKELENAEFIPMPDSPAPLSAAFSQSEKD 60

Qy        842 TLPYEELQGLKVASEAPLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVEVER 901
              |||||||||||||| | | :    |  |||:||| | ||   ||  |||||:|||||:|
Db         61 TLPYEELQGLKVASEVPPEPQRAAGAWPPRLDPAVPCVGKA--LTSAACACSALQVEVDR 118

Qy        902 LQGLVLKCLAEQQKLQQENLQIFTQLQKLNKKLEGGQQVGMHSKGTQTAKEEMEMDPKPD 961
              || |||||| ||||||||||:||||||||||||||||||||:|||||||||||||||||
Db        119 LQALVLKCLEEQQKLQQENLQMFTQLQKLNKKLEGGQQVGMHSRGTQTAKEEMEMDPKPD 178

Qy        962 LDSDSWCLLGTDSCRPSL 979
              |||:|||||||||||||
Db        179 LDSESWCLLGTDSCRPSL 196


RESULT 9
Q9VC32
ID   Q9VC32      PRELIMINARY;      PRT;   841 AA.
AC   Q9VC32;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG10951 protein.
GN   CG10951.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
```

```
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AE003749; AAF56344.1; -.
DR   FlyBase; FBgn0039230; CG10951.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00415; RCC1; 1.
DR   PRINTS; PR00633; RCCNDNSATION.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50012; RCC1_3; 4.
KW   ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   841 AA;  91814 MW;  F7C185AE4E6D69D9 CRC64;

  Query Match             16.4%;  Score 848;  DB 5;  Length 841;
  Best Local Similarity   31.2%;  Pred. No. 2.4e-55;
  Matches  223;  Conservative  125;  Mismatches  267;  Indels  100;  Gaps   21;

Qy      *    38 PR-AGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDA 96
               ||  | | |  |  :|  :||:|:|:|| | |||  |   :|:|:::|: ||   |  |
Db          90 PRETGVGVAGPELANYEKVRVVGQGSFGIAILYRRKSDGHQIVFKQINLSELSPPGRDLA 149

Qy          97 LNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKL-FEEEMVVW 155
               :||: : : | | ||::|   |: : ||||:|| :|| |   |  :: || | |  ::
Db         150 MNEVDVFSKLHHPNIVSYLGSFIKDNTLLIEMEYADGGTLAQIIAERQGKLHFPERYIIA 209

Qy         156 YLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTP 215
                  || ||:: :|   |||||:|| |:|| :  ::|:||:|::| :|::   |:|::|||
Db         210 VFEQISSAINYMHSENILHRDLKTANVFLNRRGIVKIGDFGISKIMNTKIH-AQTVLGTP 268

Qy         216 YYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSS 275
               || |||:|:| :|: |||||:||:: |:  ||:|| |:|   |  ||: |     | |
Db         269 YYFSPEMCEGKEYDNKSDIWALGCILGEMCCLKKTFAASNLSELVTKIMAG-NYTPVPSG 327

Qy         276 QYSLELIQMVHSCLDQDPEQRPTADE-------LLDRPLLRKRRREMEEKV-----TLLN 323
                |:  |  :: : |  :  :|||| |       |: | | : :    |: |       |
Db         328 -YTSGLRSLMSNLLQVEAPRRPTASEVLVYWIPLIFRSLGKNKGYSYEDDVGGPGSDQLT 386

Qy         324 APTKRPRSSTV-------------------TEAPIAVVTSRTSEVYVWGGGK------- 356
               ||      | |                   | ||  ::  |:    :   |
Db         387 APVPAAAYSNVSMELELPTAQTETKQLMIADTAAPHEILEKRSVLYQLKAFGTCFSMAPI 446

Qy         357 STPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQ 416
                 | |  ::        |   ::|| ||  :   | |     |   |||||    :::
Db         447 QLPPKAVIV-------DVAMSDSHFVVVNEDGSAYAW-----GEGTHGQLGLTALEAWKH 494

Qy         417 -PKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLN 475
                |  :| ::   :     || ||: ||  | | : ||: :  :| |:        |  :
Db         495 YPSRMESVRNYHVVSACAGDGFTILVTQAGSLLSCGSNAHLALGQDEQRNYH--SPKLIA 552

Qy         476 FFLSNPVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIV 535
                   ||||: |  ||: |:|   || ||    | |||   :|   ||   |: :::
Db         553 RLADVRVEQVAAGLQHVLALSREGAVYVWGTSTCGALGLG---NYQQQQK--FPQKILLS 607

Qy         536 AVQ-------CGCDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSF 588
                |:       || | : :|  :|::  || |::|||  :                :
Db         608 HVKTKPSKIYCGPDTSAVLFANGELHVCGSNDYNKLGFQR---------------SAKI 651

Qy         589 TLAKQLSF-YKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGP- 646
               |  |::   :|:       ||:  : | | : | | |  || |: :       | :  |
Db         652 TAFKKVQLPHKVTQACFSSTHSVFLVEGGYVYTMGRNAEGQRGIRH------CNSVDHPT 705

Qy         647 ----LGGKQVIRVSCGDEFTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICT 697
                    :  : ::: :| |: || |::|| |  .||   ||   :  |    |  |||
Db         706 LVDSVKSRYIVKANCSDQCTIVASEDNIITVWGT-RNGLPGIGSTNCGLGLQICT 759


RESULT 10
Q95XQ3
ID   Q95XQ3      PRELIMINARY;      PRT;   579 AA.
AC   Q95XQ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 65.0 kDa protein.
GN   Y39G10AR.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Bradshaw-Cordum H., Lamar B., Stoneking T.;
RT   "The sequence of C. elegans cosmid Y39G10AR.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC025716; AAK39610.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR   PROSITE; PS50012; RCC1_3; 1.
KW   ATP-binding; Hypothetical protein; Transferase.
SQ   SEQUENCE   579 AA;  64963 MW;  3B38975F5284B9AB CRC64;

  Query Match             12.9%;  Score 670.5;  DB 5;  Length 579;
  Best Local Similarity   34.9%;  Pred. No. 3.9e-42;
  Matches  160;  Conservative   84;  Mismatches  152;  Indels   63;  Gaps   12;

Qy          28 SSPGPSASQGPRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTR 87
               ::| | |:             :  |  || :|:|||| | |||| || |||: ||:::
Db         158 TNPNPRATSA----------QNRYERIRTVGKGAFGSAVLYRRREDSSLVIIKEINMYD 206

Qy          88 LSEKERRDALNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKL 147
               |   :|| |||| :|: ::| ||||||: | :   |:||:|| :|| |  ::| :   |
Db         207 LDSSQRRLALNEVSLLSRIEHPNIIAYYDSFEEEGVLMIEMEYADGGTL-AQMLSRTQNL 265

Qy         148 FEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSE--Y 205
```

```
Db        266 LDEEQIGDMMIQMCSAVAYLHENSVLHRDLKTANVFLTRDSFVKIGDFGISKIMGTETLA 325

Qy        206 SMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQ 265

Db        326 QGAKTVVGTPYYISPEMCSGVSYNEKSDMWALGCILYEMCCLKKAFEGDN-LPALVNSIM 384

Qy        266 GIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLD--RPLLRKRRR---EMEEKVT 320

Db        385 TCAYTPV-KGPYSAEMKMVIRELLQLDPQKRPSAPQALKMLRPSENRHRHTSGSMRSSVS 443

Qy        321 L-----LNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVC 375

Db        444 FTTLYDLHVP-------TITLSSIPDLPSR----------------------ISIKQVA 473

Qy        376 AGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGD 435

Db        474 LSKTHTMVLTNDNELFGF-----GDNSCGQLGFGKFSKMENGRLVESLKGRDIQSVGVGN 528

Qy        436 DFTVCVTDEGQLYAFGSDYYGCMGVD--KVAGPEVLEPM 472

Db        529 YFSVICCDRGTLLTFGTAKNLGLGKDAKDSAKPMLLEQL 567


RESULT 11
Q96PY6
ID   Q96PY6      PRELIMINARY;      PRT;  1265 AA.
AC   Q96PY6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   KIAA1901 protein (Fragment).
GN   KIAA1901.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21456161; PubMed=11572484;
RA   Nagase T., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XXI.
RT   The complete sequences of 60 new cDNA clones from brain which code for
RT   large proteins.";
RL   DNA Res. 8:179-187(2001).
DR   EMBL; AB067488; BAB67794.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
FT   NON_TER       1       1
SQ   SEQUENCE   1265 AA;  143754 MW;  31DDF2DFEAE39C89 CRC64;

  Query Match           11.6%;  Score 599.5;  DB 4;  Length 1265;
  Best Local Similarity   37.5%;  Pred. No. 3.1e-36;
  Matches  124;  Conservative   71;  Mismatches  113;  Indels   23;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111

Db         11 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 70

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171

Db         71 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 130

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231

Db        131 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 190

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291

Db        191 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 248

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335

Db        249 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 308

Qy        336 EAPIAVVTSRTSE-----YYVWGGGKSTPQK 361

Db        309 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 339


RESULT 12
O35673
ID   O35673      PRELIMINARY;      PRT;  744 AA.
AC   O35673;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   MSTK2S kinase-like protein.
GN   NEK4 OR MSTK2S.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=20001940; PubMed=10529384;
RA   Hayashi K., Igarashi H., Ogawa M., Sakaguchi N.;
RT   "Activity and substrate specificity of the murine STK2
RT   serine/threonine kinase that is structurally related to the mitotic
RT   regulator protein NIMA of Aspergillus nidulans.";
RL   Biochem. Biophys. Res. Commun. 264:449-456(1999).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; Y09234; CAA70436.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344404; Nek4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   744 AA;  83542 MW;  82E1AAF3E9DCC72A CRC64;

  Query Match           10.8%;  Score 561;  DB 11;  Length 744;
  Best Local Similarity   33.0%;  Pred. No. 1.1e-33;
  Matches  140;  Conservative   72;  Mismatches  146;  Indels   66;  Gaps   11;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111

Db          6 YCYMRVVGRGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170

Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPESQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230

Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHGDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290

Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEG--KLPPMPKVYSTELAELIRTMLS 243

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTE------APIAVVTS 344

Db        244 RRPEERPSVRSILRQPYIK-------HHISLFLEATKAKTSKNNVKNCDSRAKPVAAVVS 296
```

```
Qy        345 RTSE----------------VYVWGGGKSTPQKLDV----IKSGCSAR----QVCAGNT 379
              |  |                  ::| |  |   |:  |    ::|  |      :    ||
Db        297 RKEESNTDVIHYQPRSSEGSALHVMGEDKCLSQEKPVDIGPLRSPASLEGHTGKQDMNNT 356

Qy        380 HFAVVTVEKELYTWVNM----------QGGTKLHGQLGHGDKA-------SYRQPKHVE 421
                :  |: :     :|:             |     |  | | |       |  | |  |
Db        357 GESCATISR-----INIDILPAERRDSANAGVVQESQPQHVDAADEVDSQCSISQEK--E 409

Qy        422 KLQG 425
              :|||
Db        410 RLQG 413


RESULT 13
Q9R1J1
ID   Q9R1J1      PRELIMINARY;      PRT;   792 AA.
AC   Q9R1J1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Serine/threonine-protein kinase NEK4.
GN   NEK4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99321807; PubMed=10393247;
RA   Chen A., Yanai A., Arama E., Kilfin G., Motro B.;
RT   "NIMA-related kinases: isolation and characterization of murine nek3
RT   and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3.";
RL   Gene 234:127-137(1999).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AF099067; AAD16287.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344404; Nek4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   792 AA;  89093 MW;  9B8F373117377CC6 CRC64;

  Query Match           10.8%;  Score 561;  DB 11;  Length 792;
  Best Local Similarity   33.0%;  Pred. No. 1.2e-33;
  Matches  140;  Conservative   72;  Mismatches  146;  Indels   66;  Gaps   11;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
                |  :||:|||::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYMRVVGRGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPESQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:|||: |:||:|| |:|: | :   || ||:|||||||||     ||:
Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHGDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||:||:|| ::|: |||  |:| :  :|  :|::|    :      || || ::: : |
Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEG--KLPPMPKVYSTELAELIRTMLS 243

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTE------APIAVVTS 344
              : ||:||:   :| :| ::         ::|    ||   |    :       |:| | |
Db        244 RRPEERPSVRSILRQPYIK-------HHISLFLEATKAKTSKNNVKNCDSRAKPVAAVVS 296

Qy        345 RTSE----------------VYVWGGGKSTPQKLDV----IKSGCSAR----QVCAGNT 379
              |  |                  ::| |  |   |:  |    ::|  |      :    ||
Db        297 RKEESNTDVIHYQPRSSEGSALHVMGEDKCLSQEKPVDIGPLRSPASLEGHTGKQDMNNT 356

Qy        380 HFAVVTVEKELYTWVNM----------QGGTKLHGQLGHGDKA-------SYRQPKHVE 421
                :  |: :     :|:             |     |  | | |       |  | |  |
Db        357 GESCATISR-----INIDILPAERRDSANAGVVQESQPQHVDAADEVDSQCSISQEK--E 409

Qy        422 KLQG 425
              :|||
Db        410 RLQG 413


RESULT 14
Q9Z1J2
ID   Q9Z1J2      PRELIMINARY;      PRT;   792 AA.
AC   Q9Z1J2;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Serine/threonine kinase protein MSTK2L,long-form.
GN   NEK4 OR MSTK2L.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=SPLEEN;
RX   MEDLINE=20001940; PubMed=10529384;
RA   Hayashi K., Igarashi H., Ogawa M., Sakaguchi N.;
RT   "Activity and substrate specificity of the murine STK2
RT   serine/threonine kinase that is structurally related to the mitotic
RT   regulator protein NIMA of Aspergillus nidulans.";
RL   Biochem. Biophys. Res. Commun. 264:449-456(1999).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AJ223071; CAA11072.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344404; Nek4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   792 AA;  88994 MW;  CF9187311C807A1D CRC64;

  Query Match           10.8%;  Score 561;  DB 11;  Length 792;
  Best Local Similarity   33.0%;  Pred. No. 1.2e-33;
  Matches  140;  Conservative   72;  Mismatches  146;  Indels   66;  Gaps   11;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
                |  :||:|||::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYMRVVGRGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPESQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:|||: |:||:|| |:|: | :   || ||:|||||||||     ||:
Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHGDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||:||:|| ::|: |||  |:| :  :|  :|::|    :      || || ::: : |
Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEG--KLPPMPKVYSTELAELIRTMLS 243
```

```
Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTE------APIAVVTS 344
              : ||:||:   :| :| ::         ::|    ||   |    :       |:| | |
Db        244 RRPEERPSVRSILRQPYIK-------HHISLFLEATKAKTSKNNVKNCDSRAKPVAAVVS 296

Qy        345 RTSE----------------VYVWGGGKSTPQKLDV----IKSGCSAR----QVCAGNT 379
              |  |                  ::| |   |    |:  |    ::|  |      :    ||
Db        297 RKEESNTDVIHYQPRSSEGSALHVMGEDKCLSQEKPVDIGPLRSPASLEGHTGKQDMNNT 356

Qy        380 HFAVVTVEKELYTWVNM----------QGGTKLHGQLGHGDKA-------SYRQPKHVE 421
              :  |: :     :|:           |     |  | | |       |  | | |
Db        357 GESCATISR-----INIDILPAERRDSANAGVVQESQPQHVDAADEVDSQCSISQEK--E 409

Qy        422 KLQG 425
              :|||
Db        410 RLQG 413
```

```
RESULT 15
Q9D685
ID   Q9D685      PRELIMINARY;      PRT;   291 AA.
AC   Q9D685;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   4632401F23Rik protein.
GN   4632401F23RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SKIN;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AK014546; BAB29424.1; -.
DR   MGD; MGI:1921589; 4632401F23Rik.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   291 AA;  32802 MW;  DF2B752B4467021F CRC64;
```

```
  Query Match             10.4%;  Score 541;  DB 11;  Length 291;
  Best Local Similarity   44.7%;  Pred. No. 8.9e-33;
  Matches  115;  Conservative   54;  Mismatches   84;  Indels    4;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  |||:|||| |   |  |  |  ||: |:: : :::::||: | ||  :| || | |:
Db          4 YERIRVVGRGALGIVHLCLRKADQKLVILKQIPVEQMTKEERQAAQNECQVLKLLNHPNV 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              | || :|:::  |:| :||  || | : | :: : | ||| :: :  ||: |:  :|
Db         64 IEYYENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL 123

Qy        172 ILHRDIKTLNIFLTKANL-IKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              ||||:|| || | |  : :|:||::|::| |:|: | | |:|||| |:|||||:|  ||
Db        124 ILHRDLKTQNILLDKHRMVVKIGDFGISKILSSK-SKAYTVVGTPCYISPELCEGKPYNQ 182

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              ||||||:|||::|| :||| |:| |   | :||: |  |    | :|| || |:| | |
Db        183 KSDIWALGCVLYELASLKRAFEAANLPALVLKIMSGTFA--PISDRYSPELRQLVLSLLS 240

Qy        291 QDPEQRPTADELLDRPL 307
              :| | |    :: :||
Db        241 LEPAQGPPLSHIMAQPL 257
```

```
Search completed: July  1, 2003, 08:33:52
Job time : 47.4387 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 34.208 Seconds
(without alignments)
2751.276 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	601	11.6	774	2	S25284	protein kinase nek
2	577	11.1	841	1	I78885	serine/threonine-s
3	561	10.8	792	2	JC7122	protein kinase (EC
4	499	9.6	445	2	G01452	NIMA-like protein
5	476	9.2	699	2	A43734	probable protein k
6	461.5	8.9	431	2	T11854	protein kinase (EC
7	461.5	8.9	779	2	A57177	NIMA-like protein
8	439	8.5	722	2	T37970	probable G2-specif
9	417	8.0	440	2	T50662	UVB-resistance pro
10	412	7.9	357	2	T29771	hypothetical prote
11	412	7.9	4836	2	T14346	herc2 protein - mo
12	397	7.7	435	2	S23580	probable protein k
13	393.5	7.6	1996	2	F71405	probable TMV resis
14	387.5	7.5	294	2	T21075	hypothetical prote
15	381.5	7.4	1895	2	T15881	hypothetical prote
16	377	7.3	1080	2	S48944	hypothetical prote
17	375.5	7.2	941	2	T49136	protein kinase-lik
18	371.5	7.2	312	2	T38525	serine/threonine p
19	368	7.1	648	2	T47988	serine/threonine-p
20	363	7.0	883	2	A96662	hypothetical prote
21	361	7.0	1233	2	T14157	serine/threonine p
22	360.5	7.0	836	2	B96716	probable serine/th
23	357.5	6.9	465	2	B55748	protein kinase (EC
24	357.5	6.9	925	2	A55748	protein kinase (EC
25	357	6.9	4861	2	S71752	giant protein p619
26	356	6.9	603	2	S34130	serine/threonine-s
27	355	6.8	1054	2	B38919	hypothetical prote
28	352	6.8	682	2	A44493	serum-inducible ki
29	351	6.8	1231	2	T18532	serine/threoine pr
30	349.5	6.7	1006	2	E96683	hypothetical prote
31	349	6.7	608	2	G96575	probable MEK kinas
32	349	6.7	631	2	A57286	probable serine/th
33	349	6.7	1062	2	S46367	protein kinase CDC
34	348	6.7	576	2	S22127	protein kinase poL
35	347	6.7	603	2	A54596	protein kinase - m
36	346.5	6.7	651	2	A96591	NPK1-related prote
37	345.5	6.7	1206	2	T34021	protein kinase SK2
38	345	6.7	603	2	A47545	protein kinase (EC
39	339	6.5	200	2	B96587	hypothetical prote
40	339	6.5	690	2	C96572	protein F12M16.4 [
41	339	6.5	1142	2	S59359	GIN4 protein - yea
42	339	6.5	1518	2	S37928	probable purine nu
43	334	6.4	1051	1	JW0051	serine/threonine-s
44	333.5	6.4	1108	2	D96798	hypothetical prote
45	332	6.4	1028	2	A96719	hypothetical prote

ALIGNMENTS

RESULT 1
S25284
protein kinase nek1 (EC 2.7.1.-) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Sep-1999
C;Accession: S25284
R;Letwin, K.; Mizzen, L.; Motro, B.; Ben-David, Y.; Bernstein, A.; Pawson, T.
EMBO J. 11, 3521-3531, 1992
A;Title: A mammalian dual specificity protein kinase, Nek1, is related to the NIMA ce
A;Reference number: S25284; MUID:93010942; PMID:1382974
A;Accession: S25284
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-774 <LET>
A;Cross-references: GB:S45828; NID:g256854; PIDN:AAB23529.1; PID:g256855
C;Genetics:
A;Gene: nek1
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP; phosphotransferase
F;2-258/Domain: protein kinase homology <KIN>
F;10-18/Region: protein kinase ATP-binding motif

Query Match 11.6%; Score 601; DB 2; Length 774;
Best Local Similarity 39.8%; Pred. No. 5.6e-19;
Matches 117; Conservative 67; Mismatches 94; Indels 16; Gaps 3;

```
Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||     | ||:::|:|:|||:::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |. | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYKESFEENGSLYIVMDYCEGGDLFKRINAQKGALFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              ||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| ||| |:| | || :||: |  :    | || :|  ::     :
Db        184 SDIWALGCVLYELCTLKHAFEAGNMKNLVLKIISG--SFPPVSPHYSYDLRSLLSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRRE------MEEKVTLLN--------APTKRPRS 331
              :|  ||: : :|::  : ||  :      : |:  |           | ||| |
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTLSKFGPQPLPGKRPAS 295
```

RESULT 2
I78885
serine/threonine-specific protein kinase (EC 2.7.1.-) STK2 - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999

```
C;Accession: I78885
R;Levedakou, E.N.; He, M.; Baptist, E.W.; Craven, R.J.; Cance, W.G.; Welcsh, P.L.; Simmc
Oncogene 9, 1977-1988, 1994
A;Title: Two novel human serine/threonine kinases with homologies to the cell cycle regu
A;Reference number: I58396; MUID:94268838; PMID:8208544
A;Accession: I78885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-841 <RES>
A;Cross-references: GB:L20321; NID:g348244; PIDN:AAA36658.1; PID:g348245
C;Genetics:
A;Gene: GDB:STK2
A;Cross-references: GDB:374125
A;Map position: 3p21.1-3p21.1
C;Superfamily: human serine/threonine-specific protein kinase STK2; protein kinase homol
C;Keywords: phosphotransferase
F;4-261/Domain: protein kinase homology.<KIN>

  Query Match            11.1%;  Score 577;  DB 1;  Length 841;
  Best Local Similarity   35.4%;  Pred. No. 6.6e-18;
  Matches  128;  Conservative   71;  Mismatches  131;  Indels   32;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|:|::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYLRVVGKGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPENQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| |:|||: |:||:|| |:|: | :   || ||:||||||||||     ||:
Db         126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHCDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDSSQYSLELIQMVHSCL 289
              |||:||:|| ::|: |||  |:| :  :|  :|::| : ||  |   || || ::: : |
Db         186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEGKLPAMPRD---YSPELAELIRTML 242

Qy        290 DQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVV------- 342
              : ||:||:   :| :| : ||:     : | :       ::      | | |
Db         243 SKRPEERPSVRSILRQPYI-KRQISFFLEATKIKTSKNNIKNGDSQSKPFATVVSGEAES 301

Qy        343 ----------TSRTSEVYVWGGGK----STPQKLDVIKSGCS-----ARQVCAGNTHFAV 383
                       :|  |: |: | ||      |:   ::|| |      :|  :  |  |
Db         302 NHEVIHPQPLSSEGSQTYIMGEGKCLSQEKPRASGLLKSPASLKAHTCKQDLSNTTELAT 361

Qy        384 VT 385
              ::
Db         362 IS 363


RESULT 3
JC7122
protein kinase (EC 2.7.1.37) 2 - mouse (strain balb/c)
N;Alternate names: serine (threonine) protein kinase
C;Species: Mus musculus (house mouse)
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: JC7122
R;Hayashi, K.; Igarashi, H.; Ogawa, M.; Sakaguchi, N.
Biochem. Biophys. Res. Commun. 264, 449-456, 1999
A;Title: Activity and substrate specificity of the murine STK2 serine/threonine kinase t
A;Reference number: JC7122; MUID:20001940; PMID:10529384
A;Accession: JC7122
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-792 <HAY>
A;Cross-references: GB:AJ223071; NID:g4138208; PID:g4138209
C;Genetics:
A;Gene: MSTK2L
C;Superfamily: human serine/threonine-specific protein kinase STK2; protein kinase homol
C;Keywords: phosphotransferase

  Query Match            10.8%;  Score 561;  DB 2;  Length 792;
  Best Local Similarity   33.0%;  Pred. No. 3e-17;
  Matches  140;  Conservative   72;  Mismatches  146;  Indels   66;  Gaps   11;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|||::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYMRVVGRGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLYIVMGFCEGGDLYRKLKEQKGQLLPESQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              |||||:|| |:|||: |:||:|| |:|: | :   || ||:|||||||||||     ||:
Db         126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHGDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||:||:|| ::|: |||  |:| :  :|  :|::|   :      || || ::: : |
Db         186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEG--KLPPMPKVYSTELAELIRTMLS 243

Qy        291 QDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTE------APIAVVTS 344
              : ||:||:   :| :| ::       ::|    ||   |   :       |:| | |
Db         244 RRPEERPSVRSILRQPYIK-------HHISLFLEATKAKTSKNNVKNCDSRAKPVAAVVS 296

Qy        345 RTSE-----------------VYVWGGGKSTPQKLDV----IKSGCSAR----QVCAGNT 379
              |  |                  ::| |  |   |:  |    ::|  |       :    ||
Db         297 RKEESNTDVIHYQPRSSEGSALHVMGEDKCLSQEKPVDIGPLRSPASLEGHTGKQDMNNT 356

Qy        380 HFAVVTVEKELYTWVNM----------QGGTKLHGQLGHGDKA-------SYRQPKHVE 421
                 :  |: :     :|:            |      |  | | |       |  | |  |
Db         357 GESCATISR-----INIDILPAERRDSANAGVVQESQPQHVDAADEVDSQCSISQEK--E 409

Qy        422 KLQG 425
              :|||
Db         410 RLQG 413


RESULT 4
G01452
NIMA-like protein kinase 1 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 24-Sep-1999
C;Accession: G01452
R;Lu, K.P.
submitted to the EMBL Data Library, June 1994
A;Reference number: G07172
A;Accession: G01452
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-445 <LUX>
A;Cross-references: EMBL:U11050; NID:g507874; PIDN:AAA19558.1; PID:g507875
C;Genetics:
A;Gene: NLK1
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP
F;6-271/Domain: protein kinase homology <KIN>
F;14-22/Region: protein kinase ATP-binding motif

  Query Match             9.6%;  Score 499;  DB 2;  Length 445;
  Best Local Similarity   37.0%;  Pred. No. 7.8e-15;
  Matches  113;  Conservative   62;  Mismatches  108;  Indels   22;  Gaps    7;

Qy         45 AEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILA 104
              ||  |: |     :| |::|     ||  |  ::|||||:|   ::| |::  ::|: :|
Db          5 AEDYEVLY----TIGTGSYGRCQKIRRKSDGKILVWKELDYGSMTEAEKQMLVSEVNLLR 60

Qy        105 LLQHDNIIAYYNHFMD--NTTLLIELEYCNGGNLYDKILR--QKDKLFEEEMVVWYLFQI 160
                |:| ||: ||:  :|  |||| | :||| ||:|   | :  :: :  :|| |:  : |:
Db         61 ELKHPNIVRYYDRIIDRTNTTLYIVMEYCEGGDLASVITKGTKERQYLDEEFVLRVMTQL 120

Qy        161 VSAVSCIHKAG-----ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTP 215
                |:   |:       :||||:|  |:||     :||||:|||: || : | |:| ||||
Db         121 TLALKECHRRSDGGHTVLHRDLKPANVFLDGKQNVKLGDFGLARILNHDTSFAKTFVGTP 180
```

```
Qy        216 YYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDS 274
              ||||||    : || |||||::||:::||  |   | | :   |  || :| | :
Db        181 YYMSPEQMNRMSYNEKSDIWSLGCLLYELCALMPPFTAFSQKELAGKIREGKFRRIPY-- 238

Qy        275 SQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLR-----KRRREMEEKVTLLNAPTKRP 329
              :|| || :::   |:     ||: :|:|: ||:      ::|| :| :   |  | |
Db        239 -RYSDELNEIITRMLNLKDYHRPSVEEILENPLIADLVADEQRRNLERRGRQLGEPEKSQ 297

Qy        330 RSSTV 334
              || |
Db        298 DSSPV 302


RESULT 5
A43734
probable protein kinase nimA (EC 2.7.1.-) - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 24-Sep-1999
C;Accession: A43734; S28786
R;Osmani, S.A.; Pu, R.T.; Morris, N.R.
Cell 53, 237-244, 1988
A;Title: Mitotic induction and maintenance by overexpression of a G2-specific gene that
A;Reference number: A43734; MUID:88194523; PMID:3359487
A;Accession: A43734
A;Molecule type: DNA
A;Residues: 1-699 <OSM>
A;Cross-references: GB:M20249; NID:g168065; PIDN:AAA33316.1; PID:g168066
C;Genetics:
A;Gene: nimA
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C;Keywords: ATP; autophosphorylation; nucleus; phosphoprotein; phosphotransferase; serin
F;9-295/Domain: protein kinase homology <KIN>
F;17-25/Region: protein kinase ATP-binding motif

  Query Match             9.2%;  Score 476;  DB 2;  Length 699;
  Best Local Similarity   35.6%;  Pred. No. 1.2e-13;
  Matches  113;  Conservative   59;  Mismatches  109;  Indels   36;  Gaps    6;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :  :| |:||     :|  |  ::  ||::  ::| |||     |  ||: |:| ||
Db         11 YEVLEKIGCGSFGIIRKVKRKSDGFILCRKEINYIKMSTKEREQLTAEFNILSSLRHPNI 70

Qy        112 IAYYN--HFMDNTTLLIELEYCNGGNL--YDKILRQKDKLFEEEMVVWYLFQIVSAVSCI 167
              :|||:  |   :  | : :|||  ||:|    | |:: :|  ||: |   | |:|:|:
Db         71 VAYYHREHLKASQDLYLYMEYCGGGDLSMVIKNLKRTNKYAEEDFVWRILSQLVTALYRC 130

Qy        168 H-------------------------KAGILHRDIKTLNIFLTKANLIKLGDYGLAKK 200
              |                         :  |||||:|  ||||   | :||||:||:|
Db        131 HYGTDPAEVGSNLLGPAPKPSGLKGKQAQMTILHRDLKPENIFLGSDNTVKLGDFGLSKL 190

Qy        201 LNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLC 260
              ::| :  | | ||||:|||||:|   ||  :|||||||:::||   :  |:|   : |
Db        191 MHS-HDFASTYVGTPFYMSPEICAAEKYTLRSDIWAVGCIMYELCQREPPFNARTHIQLV 249

Qy        261 VKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVT 320
              || :|  |   |  || ||  :: |||  :|: ||    |:: |::|  |||:|  :
Db        250 QKIREGKFAPLPDF--YSSELKNVIASCLRVNPDHRPDTATLINTPVIRLMRREVE--LN 305

Qy        321 LLNAPTKRPRSSTVTEA 337
              |:   ::   :|: :|
Db        306 NLSRAARKREEATMQKA 322


RESULT 6
T11854
protein kinase (EC 2.7.1.-) - Trypanosoma brucei
C;Species: Trypanosoma brucei
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C;Accession: T11854
R;Gale, M.J.; Parsons, M.
Mol. Biochem. Parasitol. 59, 111-122, 1993
A;Title: Trypanosoma brucei gene family encoding protein kinases with catalytic domains
A;Reference number: Z17363; MUID:93295429; PMID:8515773
A;Accession: T11854
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-431 <GAL>
A;Cross-references: EMBL:L03778; NID:g162169; PID:g162170
C;Genetics:
A;Note: nrkA
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP; phosphotransferase

  Query Match             8.9%;  Score 461.5;  DB 2;  Length 431;
  Best Local Similarity   29.9%;  Pred. No. 3.1e-13;
  Matches  127;  Conservative   77;  Mismatches  152;  Indels   69;  Gaps   13;

Qy         37 GPRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDA 96
              |    ||  :     |:   ::| |::||| :    || || | | :||:::|::::| |
Db         10 GTDGSGGRCK-----YLNKGIVGLGSYGEAYVAESVEDGSLCVAKVMDLSKMSQRDKRYA 64

Qy         97 LNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKI-LRQKD--KLFEEEMV 153
              :||  ||   | ||| |     :|  ||| :|: : ||| .::| ||     : |:|
Db         65 QSEIKCLANCNHPNIIRYIEDHEENDRLLIVMEFADSGNLDEQIKLRGSGDARYFQEHEA 124

Qy        154 VWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYS--MAETL 211
              ::   |:  |:  ||   :||||||: |: ||   |:||||:| : :     |  :| |
Db        125 LFLFLQLCLALDYIHSHKMLHRDIKSANVLLTSTGLVKLGDFGFSHQYEDTVSGVVASTF 184

Qy        212 VGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME 271
              |||||::|||    :|| |:|:|::| :::|:: :|: | |:|   |  |:: |  |
Db        185 CGTPYYLAPELWNNKRYNKKADVWSLGVLLYEIMGMKKPFSASNLKGLMSKVLAGTYAPL 244

Qy        272 VDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPR- 330
              ||  :| |  ::|   |  ||  ||:  |:   | : |  :          : |  |  |
Db        245 PDS--FSSEFKRVVDGILVADPNDRPSVREIFQIPYINKGLK------LFVQALKKNERI 296

Qy        331 SSTVTEAPIAVVTSRTSEVYV----------------------------WGGGKS-TPQ 360
              | :| |    |: :: ||:                             | |||  |:
Db        297 SDSVKE----VLVTQVSEILSSEVSPDAHRFLVSQINYDVTHRGHVNKLGGGNGKSWKPR 352

Qy        361 KLDVIK--------------SGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQ- 405
              | :::               | :  ||  |       | :::    :|  ||  :  |
Db        353 FLQIVRGQLILTDDEEGNNPKGLNLEQV-QGACPVPHSTAKRDFVFALNTVGGKGMWFQA 411

Qy        406 LGHGD 410
              : |||
Db        412 VSHGD 416


RESULT 7
A57177
NIMA-like protein kinase - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 24-Sep-1999
C;Accession: A57177
R;Pu, R.T.; Xu, G.; Wu, L.; Vierula, J.; O'Donnell, K.; Ye, X.S.; Osmani, S.A.
J. Biol. Chem. 270, 18110-18116, 1995
A;Title: Isolation of a functional homolog of the cell cycle-specific NIMA protein ki
A;Reference number: A57177; MUID:95355415; PMID:7629122
A;Accession: A57177
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-779 <PUA>
A;Cross-references: GB:L42573; NID:g1040682; PIDN:AAA80145.1; PID:g1040683
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP
F;5-290/Domain: protein kinase homology <KIN>
F;13-21/Region: protein kinase ATP-binding motif

  Query Match             8.9%;  Score 461.5;  DB 2;  Length 779;
  Best Local Similarity   36.1%;  Pred. No. 5.6e-13;
  Matches  107;  Conservative   52;  Mismatches  102;  Indels   35;  Gaps    8;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :  :| |:||     ||  |  ::  ||:   ::|:|||     |  ||: |:| ||
```

```
Db          7 YELLEKIGHGSFGIIRKVRRKADGMILCRKEISYLKMSQKEREQLHAEFSILSTLRHPNI 66

Qy        112 IAYYN--HFMDNTTLLIELEYCNGGNLYDKILRQ--KDKLFEEEMVVWYLF-QIVSAVSC 166
              : ||:  |      | : :|||  |:|  :::|   |:  : ||  || :| |:|:|:
Db         67 VGYYHREHLKATQDLHLYMEYCGNGDL-GRVIRNLIKNNQYAEESFVWSIFSQLVTALYR 125

Qy        167 IH------KAG--------------------ILHRDIKTLNIFLTKANLIKLGDYGLAKK 200
               |      : |                    |||||:|  |:|| : | :||||:||:|
Db        126 CHYGVDPPEVGKTVLGLGSTARPKPPSGGMTILHRDLKPENVFLGEDNSVKLGDFGLSKV 185

Qy        201 LNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLC 260
              : | :  | | ||||:|||||:|   ||  |||||::||:|:||   :  |:|     |
Db        186 MQS-HDFASTYVGTPFYMSPEICAAEKYTLKSDIWSLGCIIYELCAREPPFNAKTHYQLV 244

Qy        261 VKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREME 316
               || :|   :    | || ||   :  ||  :|::||    ||: |::|  |:| |
Db        245 QKIKEG--KIAPLPSVYSGELFATIKDCLRVNPDRRPDTATLLNLPIVRLMRKEKE 298


RESULT 8
T37970
probable G2-specific protein kinase (EC 2.7.1.-) - fission yeast (Schizosaccharomyces pc
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T37970
R;Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21758
A;Accession: T37970
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-722 <BAD>
A;Cross-references: EMBL:Z98975; PIDN:CAB11653.1; GSPDB:GN00066; SPDB:SPAC19E9.02
A;Experimental source: strain 972h-; cosmid c19E9
C;Genetics:
A;Gene: SPDB:SPAC19E9.02
A;Map position: 1
A;Introns: 20/3; 28/3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolc
C;Keywords: phosphotransferase; protein kinase

  Query Match             8.5%;  Score 439;  DB 2;  Length 722;
  Best Local Similarity   34.4%;  Pred. No. 4.8e-12;
  Matches  101;  Conservative   60;  Mismatches  107;  Indels   26;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :  :| |:||     :| :| :|:  ||:    :: :|::   :|: ||  |:| ||
Db          4 YKILECIGHGSFGRIYKVQRLKDGALLAQKEIHFGNITRQEKQYIADEVNILRNLKHPNI 63

Qy        112 IAYYNHFMDNTTLLIEL--EYCNGGNLYDKILRQKD--KLFEEEMVVWYLFQIVSAVSCI 167
              : |    :: :  :| |  |||  |:| : | | |:  | | |: |: :  |:: |:
Db         64 VQYCGEELNRSAQVINLYMEYCGHGDLANLIQRYKEEKKRFTEQEVLKFFTQLLLALYRC 123

Qy        168 H------------------KAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMA 208
               |                  |  :||||||  |||| : | :||||:||:| |::
Db        124 HYGENAPACDSQWPREIFHPKQSVLHRDIKPANIFLDENNSVKLGDFGLSKLLDNTRVFT 183

Qy        209 ETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIR 268
              :: ||||||||||: :   |: |||:||:||||||:  |   |:  : | |   | ||
Db        184 QSYVGTPYYMSPEIIRSSPYSAKSDVWALGCVIFEICMLTHPFEGRSYLELQRNICQG-- 241

Qy        269 AMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREME-EKVTL 321
               :      || ::  ::  ||: : : |||  :||  |:|   | ::| |:| |
Db        242 NLSCWDHHYSDDVFLLIRHCLEVNSDLRPTTYQLLRSPILSDIRSKLESERVVL 295


RESULT 9
T50662
UVB-resistance protein UVR8 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50662
R;Lim, J.E.; Kliebenstein, D.J.; Landry, L.G.; Last, R.L.
submitted to the EMBL Data Library, February 1999
A;Description: Ultraviolet-B sensitive mutant of Arabidopsis is deficient in an RCC1
A;Reference number: Z25164
A;Accession: T50662
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-440 <LIM>
A;Cross-references: EMBL:AF130441; PIDN:AAD43920.1
A;Experimental source: cultivar Landsberg erecta
C;Genetics:
A;Gene: UVR8

  Query Match             8.0%;  Score 417;  DB 2;  Length 440;
  Best Local Similarity   24.6%;  Pred. No. 2.6e-11;
  Matches  124;  Conservative   76;  Mismatches  194;  Indels  110;  Gaps   12;

Qy        334 VTEAP--IAVVTSRTSE---------VYVWGGG------------KSTPQKLDVIKSGCS 370
              ||  |  : :::  |         |  || |             : :| :|  :
Db         10 VTAPPRKVLIISAGASHSVALLSGDIVCSWGRGEDGQLGHGDAEDRPSPTQLSALDGHQI 69

Qy        371 ARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQ 430
                   |  :   |      |:|:|     |    |:|||: :    |  :: | |  | |
Db         70 VSVTCGADHTVAYSQSGMEVYSW-----GWGDFGRLGHGNSSDLFTPLPIKALHGIRIKQ 124

Qy        431 VSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDN 490
              :.|||   : || ||:: ::| :  | :|:      : | | ::  |   |:: |: |
Db        125 IACGDSHCLAVTMEGEVQSWGRNQNGQLGLGDT--EDSLVPQKIQAFEGIPIKMVAAGAE 182

Qy        491 HVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQS 550
              |   :| : ::| || | || |||    |   |::|       :  | ||   |  :: |
Db        183 HTAAVTEDGDLYGWGWGRYGNLGLGDRTDRLVPERVTSTGGEKMSMVACGWRHTISVSYS 242

Qy        551 GKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTA 610
              | :  | ::: :||       :| |:              : ||   |  |: |  ||
Db        243 GALYTYGWSKYGQLGHGDLEDHLIPHK------------LEALSNSFISQISGGWRHTM 289

Qy        611 AIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATDDNH 670
              |:   |:|  :| || ||:|||    :      :  |   ::|::|||   |:| |: |:
Db        290 ALTSDGKLYGWGWNKFGQVGVGNNLDQCSPVQVRFP-DDQKVVQVSCGWRHTLAVTERNN 348

Qy        671 IFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHTILIVEKVLNS 730
              :|||| | ||:| :       |  :  ::|:
Db        349 VFAWGRGTNGQLGI-------GESVDRNFPK---------------------------- 372

Qy        731 KTIRSNSSGLSIGTVFQSSSPGGGGGGGGEEEDSQQESETPDPSGGFRGTMEADR---- 786
                          : :: |  |  |           ||   ||| |      |:|
Db        373 -------------IIEALSVDGASG--------QHIESSNIDPSSGKSWVSPAERYAVV 410

Qy        787 -GMEGLISPTEAMGNSNGASSSCP 809
                  ||   |:     ||   | |
Db        411 PDETGL---TDGSSKGNGGDISVP 431


RESULT 10
T29771
hypothetical protein ZC581.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T29771
R;Waterston, B.; Gattung, S.; Le, T.T.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid ZC581.
A;Reference number: Z20682
A;Accession: T29771
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <WAT>
A;Cross-references: EMBL:AF003134; PIDN:AAB54139.1; GSPDB:GN00019; CESP:ZC581.1
A;Experimental source: strain Bristol N2; clone ZC581
C;Genetics:
A;Gene: CESP:ZC581.1
A;Map position: 1
A;Introns: 31/3; 81/1; 120/3; 186/1; 226/3; 260/1; 288/3
```

```
C;Superfamily: Ca2+/calmodulin-dependent protein kinase I; protein kinase homology

  Query Match             7.9%;  Score 412;  DB 2;  Length 357;
  Best Local Similarity   33.2%;  Pred. No. 3.5e-11;
  Matches  109;  Conservative   65;  Mismatches  118;  Indels   36;  Gaps    9;

Qy         51 HYIPIRVLGRGAFGEATLYRRTEDDS--LVVWKEVDLTRLSEKERRDALNEIVILALLQH 108
              :|  :||:||||||   | |   | |   |: | ::   ::|||     :|: :|  :||
Db          3 NYEKVRVVGRGAFGVCWLCRGKNDASHQKVIIKLINTHGMTEKEENSIQSEVNLLKKVQH 62

Qy        109 DNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQ---KD----KLFEEEMVVWYLFQIV 161
               || | : |: :  | | ::|  || |   |  |   ||    : | |: |: |  ||:
Db         63 PLIIGYIDSFIMDNQLGIVMQYAEGGTLERLINDQRAIKDSNMREYFPEKTVLDYFTQIL 122

Qy        162 SAVSCIHKAGILHRDIKTLNIFLT-KANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSP 220
              |:: :|:  |:|||:|  || :  :  ::|| |:|:::|:| :: | | |::||| |:||
Db        123 IALNHMHQKNIVHRDLKPQNILMNRRKTVLKLSDFGISKELGTK-SAASTVIGTPNYLSP 181

Qy        221 ELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLE 280
              |:|:   || |||:|::|||::||| |:| ||  |   : :|| :  :    |    | :
Db        182 EICESRPYNQKSDMWSLGCVLYELLQLERAFDGENLPAIVMKITRSKQNPLGD--HVSND 239

Qy        281 LIQMVHSCLDQDPEQRPTADELLDRPLL------------------RKRRREMEEKVTL 321
              :  :| : |    ::||   :||  ||:                  ||:        |
Db        240 VKMLVENLLKTHTDKRPDVSQLLSDPLVLPYLISIHCDLGRIEPPPTDKRKPSASLSSRL 299

Qy        322 LNAPTK---RPRSSTVTEAPIAVVTSRT 346
                ||:   || | : : ||   :|  |
Db        300 RTYPTQSTLRPYSLS-SNAPTTHLTQLT 326


RESULT 11
T14346
herc2 protein - mouse
N;Alternate names: rjs protein
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 19-May-2000
C;Accession: T14346; T14317
R;Walkowicz, M.J.; Ji, Y.; Ren, X.; Horsthemke, B.; Francis, F.; Russell, L.B.; Johnson,
submitted to the EMBL Data Library, June 1998
A;Description: Genomic alterations within the Herc2 gene are associated with juvenile le
A;Reference number: Z17996
A;Accession: T14346
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4836 <WAL>
A;Cross-references: EMBL:AF071173; NID:g4079810; PID:g4079811; PIDN:AAD08658.1
R;Lehman, A.L.; Nakatsu, Y.; Ching, A.; Bronson, R.T.; Oakey, R.J.; Keiper-Hyrnko, N.; F
Proc. Natl. Acad. Sci. U.S.A. 95, 9436-9441, 1998
A;Title: A very large novel protein with diverse functional motifs is deficient in rjs (
A;Reference number: Z17976; MUID:98356175; PMID:9689098
A;Accession: T14317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-691,'L',693-723,'A',725-746,'R',748-755,'L',757-928,'R',930-1113,'S',1115-
-3160,'T',3162-3384,'LL',3387-3507,'A',3509-4186,'S',4188-4715,'A',4717-4722,'R',4724-47
A;Cross-references: EMBL:AF061529; NID:g3414808; PID:g3414809; PIDN:AAC31431.1
A;Experimental source: strain C57Bl/6
C;Comment: This protein has several motifs in common with human giant protein p619.
C;Genetics:
A;Gene: Herc2; rjs
A;Map position: 7
C;Superfamily: mouse herc2 protein; ubiquitin-protein ligase homology

  Query Match             7.9%;  Score 412;  DB 2;  Length 4836;
  Best Local Similarity   23.3%;  Pred. No. 4.7e-10;
  Matches  182;  Conservative  106;  Mismatches  257;  Indels  236;  Gaps   33;

Qy        262 KIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEE---- 317
              :: : || :  ||   |::::   ||   ::           |  |::   |  ::
Db       3902 EVAKKIRELMADSE--SMDVLHESHSIFKREQ----------DEQLVQWMNRRPDDWTLS 3949

Qy        318 --------------------KVTLLNAPTKRPRSSTVTEAPIAVV--------TSRTSE 348
                                  :   :  ||  |  :  |  |: ::         :   :
Db       3950 AGGSGTIYGWGHNHRGQLGGIEGAKVKVPT--PCEALATLRPVQLIGGEQTLFAVTADGK 4007

Qy        349 VYVWG---GGK---------STPQKLDVIKSGCSAR-QVCAGNTHFAVVTVEKELYTWVN 395
              :|  |   ||:          |||  |: |:      :  | :|  |   :: | |:|:|
Db       4008 LYATGYGAGGRLGIGGTESVSTPTLLESIQHVFIKKVAVNSGGKHCLALSSEGEVYSW-- 4065

Qy        396 MQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYY 455
                 |    |:||||:::   :|: :| |:|  :  |: |   : |||  | || :|   |
Db       4066 ---GEAEDGKLGHGNRSPCDRPRVIESLRGIEVVDVAAGGAHSACVTAAGDLYTWGKGRY 4122

Qy        456 GCMG---VDKVAGPEVLEPMQLNFFLSNPVEQVSC--GDNHVVVLTRNKEVYSWGCGEYG 510
               | :|    :    |:::| :|      : |  ::|  ||   : || :  |:||| |:||
Db       4123 GRLGHSDSEDQLKPKLVEALQ-----GHRVIDIACGSGDAQTLCLTDDDTVWSWGDGDYG 4177

Qy        511 RLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQCM 570
              :||    :    | |:|    | :| |:||   :  ||:|| |   |  ::::||
Db       4178 KLGRGGSDGCKVPMKIDSLTGLGVVKVECGSQFSVALTKSGAVYTWGKGDYHRLG----- 4232

Qy        571 SGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQLG 630
                | :   |        : |   |:  || |  |     | | : |:| |  ||||
Db       4233 -----HGSDDHVRRPRQ---VQGLQGKKVIAIATGSLHCVCCTEDGEVYTWGDNDEGQLG 4284

Qy        631 VGNYKKRLGINLLGGP-----LGGKQVIRVSCGDEFTIAATDDNHIFAWGN---GGNGRL 682
               |        | :  |     | ||:| ||:||           |  ||        |:|
Db       4285 DGT------TNAIQRPRLVAALQGKKVNRVACGSA---------HTLAWSTSKPASAGKL 4329

Qy        683 -AMTPTERPHGSDICTSWPRPIFG------SLHHVPDLSCR------------------- 716
               |  | |  |  :|      ||          |||: :| |
Db       4330 PAQVPMEYNHLQEI------PIIALRNRLLLLHHISELFCPCIPMFDLEGSLDETGLGPS 4383

Qy        717 -GWHT---ILIVE-------KVLNSKTIRSNSSG----LSIGTVFQSSSPGGGGGGGGGE 761
               |: |   ||| :       ||: : :|    |    |:   | :| | :|| |  | :
Db       4384 VGFDTLRGILISQGKEAAFRKVVQATMVRDRQHGPVVELNRIQVKRSRSKGGLAGPDGTK 4443

Qy        762 EEDSQ----QESETPDP--------------------SGGFRGTM-----EADRGMEGLI 792
                |       | :||                     ||:  ::     |   |:  |:
Db       4444 SVFGQMCAKMSSFSPDSLLLPHRVWKVKFVGESVDDCGGGYSESIAEICEELQNGLTPLL 4503

Qy        793 SPTEAMGNSNGASSSCPGWLRKELENAEFIPMPDS---------------PSPLS----- 832
               |    : :||:  |  :|   |  |   |:  |                 ||||
Db       4504 IVTPNGRDESGANRDC--YL---LNPATRAPVHCSMFRFLGVLLGIAIRTGSPLSLNLAE 4558

Qy        833 -------------AAFSESEKDTLP-----------YEELQGLKVASEAPLEHKPQVEAS 868
                          |  || :|| :|           || : : :    |      :: |
Db       4559 PVWKQLAGMSLTIADLSEVDKDFIPGLMYIRDNEATSEEFEAMSLPFTVPSASGQDIQLS 4618

Qy        869 S 869
               |
Db       4619 S 4619


RESULT 12
S23580
probable protein kinase KIN3 (EC 2.7.1.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: probable protein kinase FUN52; probable protein kinase NPK1; prote
C;Species: Saccharomyces cerevisiae
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 24-Sep-1999
C;Accession: S23580; S20123; S40906; S43451; S11185
R;Schweitzer, B.; Philippsen, P.
Mol. Gen. Genet. 234, 164-167, 1992
A;Title: NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae with si
A;Reference number: S23580; MUID:92357012; PMID:1495480
A;Accession: S23580
A;Molecule type: DNA
A;Residues: 1-435 <SCH>
A;Cross-references: EMBL:X60549; NID:g298024; PIDN:CAA43042.1; PID:g298026
R;Davies, C.J.; Hutchison III, C.A.
submitted to the EMBL Data Library, January 1992
A;Description: A directed DNA sequencing strategy based upon Tn3 transposon mutagenes
A;Reference number: S20123
A;Accession: S20123
A;Molecule type: DNA
```

```
A;Residues: 1-435 <DAV>
A;Cross-references: EMBL:M67445
R;Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Oue
submitted to the EMBL Data Library, November 1993
A;Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the 4
A;Reference number: S40891
A;Accession: S40906
A;Molecule type: DNA
A;Residues: 1-430 <CLA>
A;Cross-references: EMBL:L22015; MIPS:YAR018c
R;Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Oue
Yeast 10, 535-541, 1994
A;Title: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 42 kbp
A;Reference number: S43441; MUID:95028152; PMID:7941740
A;Accession: S43451
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-435 <CLW>
A;Cross-references: EMBL:L22015; NID:g1339990; PIDN:AAC04964.1; PID:g349756
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
R;Jones, D.G.L.; Rosamond, J.
Gene 90, 87-92, 1990
A;Title: Isolation of a novel protein kinase-encoding gene from yeast by oligodeoxyribon
A;Reference number: S11185; MUID:90337351; PMID:2199332
A;Accession: S11185
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-97,'L',99-120,'Q',122-129,'I',131-234,'SQIS',239-357,'R',359-369 <JON>
A;Cross-references: EMBL:M55416
A;Experimental source: strain SB303
C;Genetics:
A;Gene: SGD:KIN3; NPK1; FUN52
A;Cross-references: SGD:S0000071; MIPS:YAR018c
A;Map position: 1R
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C;Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F;23-343/Domain: protein kinase homology <KIN>
F;31-39/Region: protein kinase ATP-binding motif

  Query Match            7.7%;  Score 397;  DB 2;  Length 435;
  Best Local Similarity   30.2%;  Pred. No. 1.9e-10;
  Matches  106;  Conservative   53;  Mismatches  114;  Indels   78;  Gaps    8;

Qy         32 PSASQG--PRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLS 89
                 |  ||  ||:               |  :  :|||:||              |:|  |::    ::
Db         13 PQQQQGHPPRS----------EYQVLEEIGRGSFGSVRKVIHIPTKKLLVRKDIKYGHMN 62

Qy         90 EKERRDALNEIVILALLQHDNIIAYYNHFMD--NTTLLIELEYCNGGNLYDKI--LRQKD 145
                |||:  :  |   ||: |:|:||: :||   |       |  :  :|||: |:|    |      :|:
Db         63 SKERQQLIAECSILSQLKHENIVEFYNWDFDEQKEVLYLYMEYCSRGDLSQMIKHYKQEH 122

Qy        146 KLFEEEMVVWYLFQIVSAVSCIH--------------------KAGILHRDIKTLNIFL 184
               |   |::|    | |:::|:    |                        |   ::|||:|   ||||
Db        123 KYIPEKIVWGILAQLLTALYKCHYGVELPTLTTIYDRMKPPVKGKNIVIHRDLKPGNIFL 182

Qy        185 T-------------------------KAN------------LIKLGDYGLAKKLNSEY 205
               :                          : |              ::||||:||||  |  :
Db        183 SYDDSDYNINEQVDGHEEVNSNYYRDHRVNSGKRGSPMDYSQVVVKLGDFGLAKSLETSI 242

Qy        206 SMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQ 265
                  |   |  ||||||||:         |:  ||||:||||||:  :|     |   |   |   |   |   ||
Db        243 QFATTYVGTPYYMSPEVLMDQPYSPLSDIWSLGCVIFEMCSLHPPFQAKNYLELQTKIKN 302

Qy        266 GIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREME 316
               |       :            ||  |    ::||  :|  :      ||:   |||          :|    |:   ::
Db        303 G--KCDTVPEYYSRGLNAIIHSMIDVNLRTRPSTFELLQDIQIRTARKSLQ 351


RESULT 13
F71405
probable TMV resistance protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C;Accession: F71405
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: F71405
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1996 <BEV>
A;Cross-references: GB:Z97336; NID:g2244788; PID:e326886; PID:g2244793
C;Genetics:
A;Map position: 4COP9-4G3845

  Query Match            7.6%;  Score 393.5;  DB 2;  Length 1996;
  Best Local Similarity   23.9%;  Pred. No. 1.2e-09;
  Matches  183;  Conservative   88;  Mismatches  273;  Indels  223;  Gaps   31;

Qy        311 RRREMEEKVTLLNAPTK-------------------------RPRSSTVTEAPIAVVT 343
               ||      : : |:|||                                   |  ||  |   |   :
Db       1118 RRGRTSIDLGIQNSPTKFGSSDVGYERGNMLRPSTDGFRISVSSTPSCSTGTSGPDDI-- 1175

Qy        344 SRTSEVYVWGG-----------GKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYT 392
                  :|||||                    ||  |:|:       |     |      | |:|| :  |::|
Db       1176 ESLGDVYVWGEVWSDGISPDGVVNSTTVKIDI---ACGVR-------HIALVTRQGEVFT 1225

Qy        393 WVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGS 452
               |     |       |:||||  :        :||  ||  ||      |    |:||:   |    |:   |  |: :|
Db       1226 WEEEAG-----GRLGHGIQVDVCRPKLVEFLALTNIDFVACGEYHTCAVSTSGDLFTWGD 1280

Qy        453 DYY--GCMG---------VDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVVLTRNKEV 501
                 :   | :|              :|:||   :|  :|                |   |:||   |   :  |  |  ::
Db       1281 GIHNVGLLGHGSDLSHWIPKRVSGP--VEGLQ---------VLSVACGTWHSALATANGKL 1330

Qy        502 YSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCGCDGTFLL----------TQSG 551
               :::|  |  :|  ||         |         |::|  :      |   :  |||      |   :              | |
Db       1331 FTFGDGAFGVLGHGDRESVSYPKEVKMLSGLKTLKVACGVWHTVAIVEVMNQTGTSTSSR 1390

Qy        552 KVLACGLNEFNKLG---------LNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIRTIA 603
               |:     |    :  :|||          |:  | :|:                                |      ||
Db       1391 KLFTWGDGDKNRLGHGNKETYLLPTCVSSLID---------------------YNFNQIA 1429

Qy        604 PGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDEFTI 663
                 |  |  |  |':      |   :   |    |        ||||    |      :|        |:     |  |:  |    :||||
Db       1430 CGHTFTVALTTSGHVFTMGGTSHGQLGSSNSDGKLPC-LVQDRLVGEFVEEISCGDHHVA 1488

Qy        664 AATDDNHIFAWGNGGNGRLA-------MTPT------ERPHGSDICTSWPRPIFGSLHHV 710
                 |    :  :|  ||  |   ||||              |||            ||                        ||
Db       1489 VLTSRSEVFTWGKGSNGRLGHGDKDDRKTPTLVEALRER------------------HV 1529

Qy        711 PDLSCRGWHTILI-VEKVLNSKTIRSNSSGLSIGTVFQSSS-----------PGG------ 753
                :||        ::          :                     |   |                 :             ||                    ||
Db       1530 KSISCGADQSVCSGCRQAFGFTRKRHNCYNCGLVHCHACSSKKALKAALAPTPGKPHRVC 1589

Qy        754 -------GGGGGGGEEEDSQQESETPDPSGGFRGTMEADRGMEG---LISP-TEAMGNSN 802
                         |         |  ·    |       :   |   :   |   |   |       ||     ||   :        |:||  ||  :   |
Db       1590 DACYTKLKAGESGYNSNVANRNSTTPTRS--LDGTGRPDRDIRSSRILLSPKTEPVKYSE 1647

Qy        803 GASSSCPGWLRK-----ELENAEFIPMPDSPSPLSAAFSESEKDTLPYEELQGLKVASEA 857
                 ||'       :   :            |:          :     |   |   |   |   :   ||.
Db       1648 VRSSRSESSIVRASQVPALQQLRDVAFPSSLSAIQNAF--------------------- 1685

Qy        858 PLEHKPQVEASSPRLNPAVTCAGKGTPLTPPACACSSLQVE-VERLQGLVLKCLAEQQKL 916
                          ||     :|:     |                :       :|      |   :          |           ::  :::      |::    |          |     ||
Db       1686 ----KPVASSSTSTLPSGTRSSRISSP--PRSSGFSRGMIDTLKKSNGVINK---EMTKL 1736

Qy        917 QQE--NLQ-----IFTQLQKLNKKLEGGQQVGM-HSKGTQTAKEEME 955
               |  :    ||:                 |::|:|  |                :  :  ||          :   |   |   |   |:
```

```
Db       1737 QSQIKNLKEKCDNQGTEIQRLKKTAREASDLAVKHSSKHKAATEVMK 1783
```

RESULT 14
T21075
hypothetical protein F19H6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T21075; T21124
R;McMurray, A.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z19368
A;Accession: T21075
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WIL>
A;Cross-references: EMBL:Z50873; PIDN:CAA90762.1; GSPDB:GN00028; CESP:F19H6.1
A;Experimental source: clone F17E5
R;McMurray, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19378
A;Accession: T21124
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-294 <WI2>
A;Cross-references: EMBL:Z68115; PIDN:CAA92169.1; GSPDB:GN00028; CESP:F19H6.1
A;Experimental source: clone F19H6
C;Genetics:
A;Gene: CESP:F19H6.1
A;Map position: X
A;Introns: 55/3; 85/2; 113/3; 150/1; 192/3; 266/2
C;Superfamily: kinase-related transforming protein; protein kinase homology

```
  Query Match             7.5%;  Score 387.5;  DB 2;  Length 294;
  Best Local Similarity   38.7%;  Pred. No. 3.2e-10;
  Matches   84;  Conservative   43;  Mismatches   85;  Indels    5;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRL-SEKERRDALNEIVILALLQHDN 110
              :|  : :|:| | |    : |  |  |  |:: :  :  :| |:| | || :|  | | |
Db         23 FIIEKKIGKGQFSEVFRAQCTWVDLHVALKKIQVFEMVDQKARQDCLKEIDLLKQLNHVN 82

Qy        111 IIAYYNHFMDNTTLLIELEYCNGGNLYDKI--LRQKDKLFEEEMVVWYLFQIVSAVSCIH 168
              :| ||  |:|| | | ||    |::  |  ::  :|  |: :  |  |:  |:: :|
Db         83 VIRYYASFIDNNQLNIVLELAEAGDMSRMIKHFKKGGRLIPEKTIWKYFVQLARALAHMH 142

Qy        169 KAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKY 228
                |:|||| |:|:|   ::|||| || :  :|: : | :||||||||||  |   |
Db        143 SKRIMHRDIKPANVFITGNGIVKLGDLGLGRFFSSKTTAAHSLVGTPYYMSPERIQESGY 202

Qy        229 NFKSDIWAVGCVIFELLTLKRTF--DATNPLNLCVKI 263
              |||||:|: ||:::|:  |:  |  |  |  :|| ||
Db        203 NFKSDLWSTGCLLYEMAALQSPFYGDKMNLYSLCKKI 239

RESULT 15
T15881
hypothetical protein D1044.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15881
R;Pauley, A.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of C. elegans cosmid D1044.
A;Reference number: Z18423
A;Accession: T15881
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1895 <PAU>
A;Cross-references: EMBL:U00065; NID:g495681; PID:g495684; PIDN:AAA50735.1; CESP:D1044.3
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:D1044.3
A;Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2; 10

```
  Query Match             7.4%;  Score 381.5;  DB 2;  Length 1895;
  Best Local Similarity   23.6%;  Pred. No. 3.8e-09;
  Matches  162;  Conservative  117;  Mismatches  237;  Indels  169;  Gaps   29;

Qy         39 RAGGGAAEQEELHYIPIRVLGRGAFG-EATLYRRTEDDS-----LVVWKEVDLTRLSEKE 92
              :| |   | | |       || ||||   |: :: :  |     |:  ||: :| |:::|
Db        423 QAVGEVGEYELLDQ-----LGAGAFGCVYTVRKKAQSHSENPAKLLALKEIFMTNLNDRE 477

Qy         93 R----RDALNEI-VILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKL 147
                    | ::|: :|   |:| ||: |   |::|  | | ::   | :| | |:  |:|
Db        478 SDKSFGDMISEVKIIKQQLRHPNIVRYRRIFVENHRLYIVMDLIQGCSLRDLIITMKEKK 537

Qy        148 --FEEEMVVWYLFQIVSAVSCIHK-AGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSE 204
               |||: :   : |:: |:  :||   |:|||:|  || :|    : : |:||||:   |
Db        538 GNFEEKKIWAMVVQMMLALRYLHKEKQIVHRDLKPNNIMMTTDERVVITDFGLAKQKGPE 597

Qy        205 YSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIV 264
              |   ::  ||  |  ||: | : |  |:|||: || |:|:  |:  | :|| | | ::||
Db        598 Y--LKSAAGTIIYSCPEIVQNLPYGEKADIWSFGCCIYEMCQLQPVFHSTNMLTLAMQIV 655

Qy        265 QGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRP-----------TADELLDRPLLRKRRR 313
              :     :  :  :| :|  :: ||| || ||              | ||  : |::
Db        656 EA--KYDPLNEMWSDDLRFLITSCLAPDPSARPDILKVSGMCGVRLLEYLD-DVARQQAS 712

Qy        314 -------------EMEEKVTLLNAPT---KRPRSSTVT---------------------- 335
                          :::|  : ||: |   |||  |: |
Db        713 TSDMTASQSSYNIKIDESPSSLNSSTSSYKRPGRSSKTSGSGKLPPINPAPRRNHSMSAG 772

Qy        336 ----------------------EAPIAVVTSR----TSEVYVWGGGKSTPQKLDVIKSG 368
                                   || |: : |     |  :      ||  |:     |
Db        773 ETPRPSSIVCLPRITDKYSVMFPSAPSAIPSRRRVQTCSTEHPARSSSSTELKVSKQSDG 832

Qy        369 CSA-----RQV------CAGNTHFAVVTVEKELYTWVNMQGGTKLHGQ-----------L 406
              :      ||:           |  :|  :||  :       | :: |           |
Db        833 LTVSSNVLRQIQDPVLTILNQIHRILVVTDKETIS-------TSMNHQRRLVEMFRKNLL 885

Qy        407 GHGDKASYRQPKHVEKLQGKAIHQVSCG---DDFTVCVTDEGQLYAFGSDYYGCMGVDKV 463
              |  : |  :   |: ||  ::  ::       ||  |  :  :  :  |      | |:
Db        886 GRENDA-VQMKTHLRKLAAESPEEIQMNLGFSDFRP-VLVQSHINGYQKD----QKVTKI 939

Qy        464 AGPEVLEPMQLNFFLSNP-VEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYT 522
              ::   ::       ||  : |       ||:| |:  |     |:|  : : :   |
Db        940 TYEQLSACIEC-LIAENPAAKHVPHRTRAVVILRRDLLV----LGQYVNMLVPTITTYVV 994

Qy        523 PQKVDVPKALIIV--AVQCGCDGTFLLTQSGKVLACGL---------------------N 559
               ::: |  | |:|    :|| : :  :: ||:|: |                       |
Db        995 IRQIHVSLAAILVYTEYECGSNSSPQVSASGQVVTCSTNTQCASGYTCNNGACCPNTNSN 1054

Qy        560 EFNKLGLNQCMSG--IINHEAYHEV 582
              :  | | |::|  ::| : |: |
Db       1055 TCSSNGNNGCLAGQTMVNGQCYNSV 1079
```

Search completed: July 1, 2003, 08:32:26
Job time : 38.208 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 17.9185 Seconds
(without alignments)
2266.116 Million cell updates/sec

Title: US-09-884-001-4
Perfect score: 5183
Sequence: 1 MSVLGEYERHCDSINSDFGS.....PDLDSDSWCLLGTDSCRPSL 979

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	601	11.6	774	1	NEK1_MOUSE	P51954 mus musculu
2	599.5	11.6	1258	1	NEK1_HUMAN	Q96py6 homo sapien
3	577	11.1	841	1	NEK4_HUMAN	P51957 homo sapien
4	516	10.0	506	1	NEK3_HUMAN	P51956 homo sapien
5	499	9.6	445	1	NEK2_HUMAN	P51955 homo sapien
6	494.5	9.5	511	1	NEK3_MOUSE	Q9r0a5 mus musculu
7	493.5	9.5	443	1	NEK2_MOUSE	O35942 mus musculu
8	476	9.2	699	1	NIMA_EMENI	P11837 emericella
9	461.5	8.9	431	1	NRKA_TRYBB	Q08942 trypanosoma
10	461.5	8.9	779	1	NIM1_NEUCR	P48479 neurospora
11	454.5	8.8	431	1	NRKB_TRYBB	Q03428 trypanosoma
12	397	7.7	435	1	KIN3_YEAST	P22209 saccharomyc
13	381.5	7.4	1576	1	YLK3_CAEEL	P41951 caenorhabdi
14	377	7.3	1080	1	NRK1_YEAST	P38692 saccharomyc
15	357.5	6.9	968	1	ST10_HUMAN	O94804 homo sapien
16	356	6.9	576	1	POLO_DROME	P52304 drosophila
17	356	6.9	603	1	PLK1_HUMAN	P53350 homo sapien
18	355	6.8	1050	1	Y032_HUMAN	Q15034 homo sapien
19	354	6.8	685	1	SNK_HUMAN	Q9nyy3 homo sapien
20	352	6.8	682	1	SNK_MOUSE	P53351 mus musculu
21	351.5	6.8	646	1	CNK_HUMAN	Q9h4b4 homo sapien
22	349.5	6.7	682	1	SNK_RAT	Q9r012 rattus norv
23	349	6.7	631	1	CNK_MOUSE	Q60806 mus musculu
24	349	6.7	1062	1	CC7_SCHPO	P41892 schizosacch
25	347.5	6.7	966	1	ST10_MOUSE	O55098 mus musculu
26	347	6.7	591	1	PAK4_HUMAN	O96013 homo sapien
27	347	6.7	603	1	PLK1_MOUSE	Q07832 mus musculu
28	346	6.7	603	1	PLK1_RAT	Q62673 rattus norv
29	342	6.6	615	1	CNK_RAT	Q9r011 rattus norv
30	340.5	6.6	815	1	RPGR_HUMAN	Q92834 homo sapien
31	339.5	6.6	719	1	PAK7_HUMAN	Q9p286 homo sapien
32	339	6.5	1142	1	GIN4_YEAST	Q12263 saccharomyc
33	339	6.5	1518	1	KKK1_YEAST	P34244 saccharomyc
34	338.5	6.5	491	1	STK3_HUMAN	Q13188 homo sapien
35	336.5	6.5	487	1	STK4_HUMAN	Q13043 homo sapien
36	335	6.5	443	1	ST24_HUMAN	Q9y6e0 homo sapien
37	334	6.4	1051	1	ULK1_MOUSE	O70405 mus musculu
38	331.5	6.4	915	1	KCC4_YEAST	P25389 saccharomyc
39	329.5	6.4	426	1	ST25_HUMAN	O00506 homo sapien
40	329	6.3	1050	1	ULK1_HUMAN	O75385 homo sapien
41	328.5	6.3	421	1	RCC1_HUMAN	P18754 homo sapien
42	328.5	6.3	974	1	CC15_YEAST	P27636 saccharomyc
43	327.5	6.3	421	1	RCC1_MESAU	P23800 mesocricetu
44	327.5	6.3	1116	1	MKH1_SCHPO	Q10407 schizosacch
45	327	6.3	648	1	PLK1_CAEEL	P34331 caenorhabdi

ALIGNMENTS

RESULT 1
NEK1_MOUSE
ID NEK1_MOUSE STANDARD; PRT; 774 AA.
AC P51954;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase NEK1 (EC 2.7.1.37) (NimA-related
DE protein kinase 1).
GN NEK1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93010942; PubMed=1382974;
RA Letwin K., Mizzen L., Motro B., Ben-David Y., Bernstein A.,
RA Pawson T.;
RT "A mammalian dual specificity protein kinase, Nek1, is related to the
RT NIMA cell cycle regulator and highly expressed in meiotic germ
RT cells.";
RL EMBO J. 11:3521-3531(1992).
CC -!- FUNCTION: PHOSPHORYLATES SERINES AND THREONINES, BUT ALSO APPEARS
CC TO POSSESS TYROSINE KINASE ACTIVITY. IMPLICATED IN THE CONTROL OF
CC MEIOSIS.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY IN TESTES (GERM CELLS AND
CC SERTOLI CELLS). LOWER LEVELS IN OVARY (OVOCYTES AND GRANULOSA
CC CELLS), THYMUS, AND LUNG.
CC -!- DEVELOPMENTAL STAGE: IN FEMALE, EXPRESSED AS FOLLICLES ENTER THE
CC SECONDARY STAGE UNTIL OVULATION OCCURS. IN THE MALE REPRODUCTIVE
CC SYSTEM, THE EXPRESSION IS LIMITED TO SPERMATOCYTES AND SPERMATIDS.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC NIMA SUBFAMILY.
CC ----------

CC ----------
DR EMBL; S45828; AAB23529.1; -.
DR HSSP; Q63450; 1A06.
DR MGD; MGI:97303; Nek1.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

```
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division;
KW   Tyrosine-protein kinase.
FT   DOMAIN        4    258       PROTEIN KINASE.
FT   NP_BIND      10     18       ATP (BY SIMILARITY).
FT   BINDING      33     33       ATP (BY SIMILARITY).
FT   ACT_SITE    128    128       BY SIMILARITY.
SQ   SEQUENCE   774 AA;  88427 MW;  7EDE5881ACDC06FA CRC64;

  Query Match             11.6%;  Score 601;  DB 1;  Length 774;
  Best Local Similarity   39.8%;  Pred. No. 3.3e-30;
  Matches  117;  Conservative   67;  Mismatches   94;  Indels   16;  Gaps    3;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | : |||    | ||:::::|:|:|||:::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAVLVKSTEDGRHYVIKEINISRMSDKERQESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYKESFEENGSLYIVMDYCEGGDLFKRINAQKGALFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| |  || :||: |  :    |  || :|  ::     :
Db        184 SDIWALGCVLYELCTLKHAFEAGNMKNLVLKIISG--SFPPVSPHYSYDLRSLLSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRRE------MEEKVTLLN--------APTKRPRS 331
              :|  ||: : :|::  : ||  :      : |:  |           | ||| |
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTLSKFGPQPLPGKRPAS 295


RESULT 2
NEK1_HUMAN
ID   NEK1_HUMAN     STANDARD;      PRT;  1258 AA.
AC   Q96PY6; Q9Y594;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase NEK1 (EC 2.7.1.37) (NimA-related
DE   protein kinase 1) (NY-REN-55 antigen).
GN   NEK1 OR KIAA1901.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=21456161; PubMed=11572484;
RA   Nagase T., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XXI.
RT   The complete sequences of 60 new cDNA clones from brain which code for
RT   large proteins.";
RL   DNA Res. 8:179-187(2001).
RN   [2]
RP   SEQUENCE OF 444-1258 FROM N.A. (ISOFORM 2).
RC   TISSUE=Renal cell carcinoma;
RX   MEDLINE=99438124; PubMed=10508479;
RA   Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA   Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA   Old L.J.;
RT   "Antigens recognized by autologous antibody in patients with
RT   renal-cell carcinoma.";
RL   Int. J. Cancer 83:456-464(1999).
CC   -!- FUNCTION: PHOSPHORYLATES SERINES AND THREONINES, BUT ALSO APPEARS
CC       TO POSSESS TYROSINE KINASE ACTIVITY. IMPLICATED IN THE CONTROL OF
CC       MEIOSIS (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC       produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB067488; BAB67794.1; ALT_INIT.
DR   EMBL; AF155113; AAD42879.1; -.
DR   Genew; HGNC:7744; NEK1.
DR   MIM; 604588; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division;
KW   Tyrosine-protein kinase; Alternative splicing.
FT   DOMAIN        4    258       PROTEIN KINASE.
FT   NP_BIND      10     18       ATP (BY SIMILARITY).
FT   BINDING      33     33       ATP (BY SIMILARITY).
FT   ACT_SITE    128    128       BY SIMILARITY.
FT   VARSPLIC    478    521       MISSING (IN ISOFORM 2).
FT   CONFLICT   1232   1232       G -> E (IN REF. 2).
SQ   SEQUENCE   1258 AA;  142828 MW;  339C4BFA56612530 CRC64;

  Query Match             11.6%;  Score 599.5;  DB 1;  Length 1258;
  Best Local Similarity   37.5%;  Pred. No. 7.9e-30;
  Matches  124;  Conservative   71;  Mismatches  113;  Indels   23;  Gaps    5;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :: :| |:||:| | :.|||    | ||:::::|:| ||| ::  |: :|| ::| ||
Db          4 YVRLQKIGEGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNI 63

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              : |   | :| :| | ::|| ||:|: :|  ||  ||:|: :: :  ||  |:  :|
Db         64 VQYRESFEENGSLYIVMDYCEGGDLFKRINAQKGVLFQEDQILDWFVQICLALKHVHDRK 123

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              |||||||: ||||||   ::|||:|:|: |||   :| | :|||||:|||:|:   || |
Db        124 ILHRDIKSQNIFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNK 183

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              |||||:|||::|| |||  |:| :  || :||: |  :    |  || :|  :|     :
Db        184 SDIWALGCVLYELCTLKHAFEAGSMKNLVLKIISG--SFPPVSLHYSYDLRSLVSQLFKR 241

Qy        292 DPEQRPTADELLDRPLLRKRRREM-------EE-------KVTLLNAPTKRPRS--STVT 335
              :|  ||: : :|::  : ||  :        ||       |     | ||| |  ::::
Db        242 NPRDRPSVNSILEKGFIAKRIEKFLSPQLIAEEFCLKTFSKFGSQPIPAKRPASGQNSIS 301

Qy        336 EAPIAVVTSRTSE-----VYVWGGGKSTPQK 361
                |   :|   ::      |   | |   :|
Db        302 VMPAQKITKPAAKYGIPLAYKKYGDKKLHEK 332


RESULT 3
NEK4_HUMAN
ID   NEK4_HUMAN     STANDARD;      PRT;   841 AA.
AC   P51957;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase NEK4 (EC 2.7.1.37) (NimA-related
DE   protein kinase 4) (Serine/threonine protein kinase 2)
```

```
DE   (Serine/threonine-protein kinase NRK2).
GN   NEK4 OR STK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast;
RX   MEDLINE=94268838; PubMed=8208544;
RA   Levedakou E.N., He M., Baptist E.W., Craven R.J., Cance W.G.,
RA   Welcsh P.L., Simmons A., Naylor S.L., Leach R.J., Lewis T.B.,
RA   Bowcock A., Liu E.T.;
RT   "Two novel human serine/threonine kinases with homologies to the cell
RT   cycle regulating Xenopus MO15, and NIMA kinases: cloning and
RT   characterization of their expression pattern.";
RL   Oncogene 9:1977-1988(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN ADULT HEART, FOLLOWED BY
CC       PANCREAS, SKELETAL MUSCLE, BRAIN, LIVER, KIDNEY, LUNG AND
CC       PLACENTA. PRESENT IN MOST PRIMARY CARCINOMAS.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L20321; AAA36658.1; -.
DR   HSSP; Q63450; 1A06.
DR   Genew; HGNC:11399; NEK4.
DR   MIM; 601959; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation.
FT   DOMAIN        6    261       PROTEIN KINASE.
FT   NP_BIND      12     20       ATP (BY SIMILARITY).
FT   BINDING      35     35       ATP (BY SIMILARITY).
FT   ACT_SITE    131    131       BY SIMILARITY.
FT   MOD_RES     165    165       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   841 AA;  94571 MW;  0DD31920DDE7EA58 CRC64;

  Query Match           11.1%;  Score 577;  DB 1;  Length 841;
  Best Local Similarity  35.4%;  Pred. No. 1.2e-28;
  Matches  128;  Conservative   71;  Mismatches  131;  Indels   32;  Gaps    7;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :||:|:|::|| || :   |    | |:::|   | :||| |  |  :|: |:| ||
Db          6 YCYLRVVGKGSYGEVTLVKHRRDGKQYVIKKLNLRNASSRERRAAEQEAQLLSQLKHPNI 65

Qy        112 IAYYNHFMDNTTLL-IELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              : |   :     || | : :| ||:|| |:  || :|  |  || :  ||  |:  :|:
Db         66 VTYKESWEGGDGLLLYIVMGFCEGGDLYRKLKEQKGQLLPENQVVEWFVQIAMALQYLHEK 125

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
               |||||:|| |:|||: |:||:|| |:|: | :   || ||:|||||||||      ||:
Db        126 HILHRDLKTQNVFLTRTNIIKVGDLGIARVLENHCDMASTLIGTPYYMSPELFSNKPYNY 185

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDSSQYSLELIQMVHSCL 289
              |||:||:|| ::|: |||  |:| :  :|  :|::| : ||  |   || || ::: : |
Db        186 KSDVWALGCCVYEMATLKHAFNAKDMNSLVYRIIEGKLPAMPRD---YSPELAELIRTML 242

Qy        290 DQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVV------- 342
               : ||:||:   :| :| : ||:     : | :       ::      | | |
Db        243 SKRPEERPSVRSILRQPYI-KRQISFFLEATKIKTSKNNIKNGDSQSKPFATVVSGEAES 301

Qy        343 ----------TSRTSEVYVWGGGK----STPQKLDVIKSGCS-----ARQVCAGNTHFAV 383
                        :|  |: |: | ||      |:   ::||  |      :|  :  |  |
Db        302 NHEVIHPQPLSSEGSQTYIMGEGKCLSQEKPRASGLLKSPASLKAHTCKQDLSNTTELAT 361

Qy        384 VT 385
               ::
Db        362 IS 363


RESULT 4
NEK3_HUMAN
ID   NEK3_HUMAN     STANDARD;      PRT;   506 AA.
AC   P51956; Q8WUN5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase NEK3 (EC 2.7.1.37) (NimA-related
DE   protein kinase 3) (HSPK 36).
GN   NEK3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 48-506 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94368699; PubMed=7522034;
RA   Schultz S.J., Fry A.M., Suetterlin C., Ried T., Nigg E.A.;
RT   "Cell cycle-dependent expression of Nek2, a novel human protein
RT   kinase related to the NIMA mitotic regulator of Aspergillus
RT   nidulans.";
RL   Cell Growth Differ. 5:625-635(1994).
RN   [3]
RP   SEQUENCE OF 76-189 FROM N.A.
RX   MEDLINE=94100173; PubMed=8274451;
RA   Schultz S.J., Nigg E.A.;
RT   "Identification of 21 novel human protein kinases, including 3
RT   members of a family related to the cell cycle regulator nimA of
RT   Aspergillus nidulans.";
RL   Cell Growth Differ. 4:821-830(1993).
CC   -!- FUNCTION: KINASE THAT MAY PLAY A ROLE IN MITOTIC REGULATION.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- DISEASE: MIGHT BE CANDIDATE FOR USHER SYNDROME, BECAUSE OF ITS
CC       CHROMOSOMAL LOCATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC019916; AAH19916.1; ALT_INIT.
DR   EMBL; Z29067; CAA82310.1; -.
DR   EMBL; Z25434; CAA80921.1; -.
DR   HSSP; Q00534; 1BI7.
DR   Genew; HGNC:7746; NEK3.
DR   MIM; 604044; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
```

```
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division.
FT   DOMAIN        4    257       PROTEIN KINASE.
FT   NP_BIND      10     18       ATP (BY SIMILARITY).
FT   BINDING      33     33       ATP (BY SIMILARITY).
FT   ACT_SITE    127    127       BY SIMILARITY.
FT   MOD_RES     165    165       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT     54     54       L -> S (IN REF. 2).
FT   CONFLICT     77     78       IV -> LY (IN REF. 3).
FT   CONFLICT    187    189       SLG -> PSV (IN REF. 3).
SQ   SEQUENCE   506 AA;  57704 MW;  4437EB4A41A44777 CRC64;

  Query Match             10.0%;  Score 516;  DB 1;  Length 506;
  Best Local Similarity   37.6%;  Pred. No. 4e-25;
  Matches  112;  Conservative   63;  Mismatches  105;  Indels   18;  Gaps    4;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |: :|::| |:|| | | :    : :   ||: | : |    :::  | |:|| ::| ||
Db          4 YMVLRMIGEGSFGRALLVQHESSNQMFAMKEIRLPK-SFSNTQNSRKEAVLLAKMKHPNI 62

Qy        112 IAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAG 171
              :|:   |     | | :|||:||:|  || :|| ||| |:|:: :  |:   |: |||
Db         63 VAFKESFEAEGHLYIVMEYCDGGDLMQKIKQQKGKLFPEDMILNWFTQMCLGVNHIHKKR 122

Qy        172 ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFK 231
              :|||||: |||||:   :||||:| |: |::  : | | |||||: ||: : : || |
Db        123 VLHRDIKSKNIFLTQNGKVKLGDFGSARLLSNPMAFACTYVGTPYYVPPEIWENLPYNNK 182

Qy        232 SDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQ 291
              ||||::||:::|| ||| | | :  || :|: ||   :    | || ||  :|    :
Db        183 SDIWSLGCILYELCTLKHPFQANSWKNLILKVCQG--CISPLPSHYSYELQFLVKQMFKR 240

Qy        292 DPEQRPTADELLDRPLLRKRRRE--------------MEE-KVTLLNAPTKRPRSSTV 334
              :|  ||:|  || | :: :  ::             :|| | :  | | |:   | :
Db        241 NPSHRPSATTLLSRGIVARLVQKCLPPEIIMEYGEEVLEEIKNSKHNTPRKKTNPSRI 298


RESULT 5
NEK2_HUMAN
ID   NEK2_HUMAN     STANDARD;      PRT;   445 AA.
AC   P51955;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase NEK2 (EC 2.7.1.37) (NimA-related
DE   protein kinase 2) (NimA-like protein kinase 1) (HSPK 21).
GN   NEK2 OR NLK1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell, Placenta, and Nasopharynx;
RX   MEDLINE=94368699; PubMed=7522034;
RA   Schultz S.J., Fry A.M., Suetterlin C., Ried T., Nigg E.A.;
RT   "Cell cycle-dependent expression of Nek2, a novel human protein
RT   kinase related to the NIMA mitotic regulator of Aspergillus
RT   nidulans.";
RL   Cell Growth Differ. 5:625-635(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lu K.P., Hunter T.;
RT   "Molecular cloning and expression of NLK1, a human NIMA-like kinase.";
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 83-203 FROM N.A.
RX   MEDLINE=94100173; PubMed=8274451;
RA   Schultz S.J., Nigg E.A.;
RT   "Identification of 21 novel human protein kinases, including 3 members
RT   of a family related to the cell cycle regulator nimA of Aspergillus
RT   nidulans.";
RL   Cell Growth Differ. 4:821-830(1993).
CC   -!- FUNCTION: PROTEIN KINASE THAT IS INVOLVED IN MITOTIC REGULATION.
CC       MAY HAVE A ROLE AT THE G2-M TRANSITION. MAY ALSO PLAY A ROLE IN
CC       MEIOSIS.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: Interacts with TERF1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES THROUGHOUT S PHASE AND SHOWS
CC       MAXIMAL LEVELS IN LATE G2. THIS EXPRESSION PATTERN IS HIGHLY
CC       REMINISCENT OF THAT OF A AND B CYCLINS.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z29066; CAA82309.1; -.
DR   EMBL; U11050; AAA19558.1; -.
DR   EMBL; Z25425; CAA80912.1; -.
DR   HSSP; Q00534; 1BI7.
DR   Genew; HGNC:7745; NEK2.
DR   MIM; 604043; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division; Meiosis.
FT   DOMAIN        8    271       PROTEIN KINASE.
FT   NP_BIND      14     22       ATP (BY SIMILARITY).
FT   BINDING      37     37       ATP (BY SIMILARITY).
FT   ACT_SITE    141    141       BY SIMILARITY.
FT   MOD_RES     179    179       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT     84     85       IV -> LY (IN REF. 3).
SQ   SEQUENCE   445 AA;  51763 MW;  D33A37778ABB6D9E CRC64;

  Query Match              9.6%;  Score 499;  DB 1;  Length 445;
  Best Local Similarity   37.0%;  Pred. No. 3.9e-24;
  Matches  113;  Conservative   62;  Mismatches  108;  Indels   22;  Gaps    7;

Qy         45 AEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILA 104
              ||  |: |     :| |::|     ||  |  ::||||:|   ::| |::  ::|: :|
Db          5 AEDYEVLY----TIGTGSYGRCQKIRRKSDGKILVWKELDYGSMTEAEKQMLVSEVNLLR 60

Qy        105 LLQHDNIIAYYNHFMD--NTTLLIELEYCNGGNLYDKILR--QKDKLFEEEMVVWYLFQI 160
               |:| ||: ||:  :|  |||| | :||| ||:|   | :  :: :  :|| |:  : |:
Db         61 ELKHPNIVRYYDRIIDRTNTTLYIVMEYCEGGDLASVITKGTKERQYLDEEFVLRVMTQL 120

Qy        161 VSAVSCIHKAG-----ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTP 215
               |:   |:       :||||:|  |:||     :||||:|||: || : | |:| ||||
Db        121 TLALKECHRRSDGGHTVLHRDLKPANVFLDGKQNVKLGDFGLARILNHDTSFAKTFVGTP 180

Qy        216 YYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQG-IRAMEVDS 274
              ||||||    : || |||||::||:::||  |   | | :   |  || :|  | :
Db        181 YYMSPEQMNRMSYNEKSDIWSLGCLLYELCALMPPFTAFSQKELAGKIREGKFRRIPY-- 238

Qy        275 SQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLR-----KRRREMEEKVTLLNAPTKRP 329
              :|| || :::   |:     ||: :|:|: ||:       ::|| :| :   |  | |
Db        239 -RYSDELNEIITRMLNLKDYHRPSVEEILENPLIADLVADEQRRNLERRGRQLGEPEKSQ 297
```

```
Qy        330 RSSTV 334
              || |
Db        298 DSSPV 302


RESULT 6
NEK3_MOUSE
ID   NEK3_MOUSE     STANDARD;      PRT;   511 AA.
AC   Q9R0A5; Q9Z0X9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine-protein kinase NEK3 (EC 2.7.1.37) (NimA-related
DE   protein kinase 3).
GN   NEK3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99240743; PubMed=10224116;
RA   Tanaka K., Nigg E.A.;
RT   "Cloning and characterization of the murine Nek3 protein kinase, a
RT   novel member of the NIMA family of putative cell cycle regulators.";
RL   J. Biol. Chem. 274:13491-13497(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Chen A., Yanai A., Arama E., Kilfin G., Motro B.;
RT   "NIMA-related kinases: isolation and characterization of murine nek3
RT   and nek4 cDNAs.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KINASE THAT MAY PLAY A ROLE IN MITOTIC REGULATION.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF093416; AAD20986.1; -.
DR   EMBL; AF099066; AAD16286.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344371; Nek3.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division.
FT   DOMAIN        4    255       PROTEIN KINASE.
FT   NP_BIND      10     18       ATP (BY SIMILARITY).
FT   BINDING      33     33       ATP (BY SIMILARITY).
FT   ACT_SITE    125    125       BY SIMILARITY.
FT   MOD_RES     163    163       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT    239    239       N -> K (IN REF. 2).
FT   CONFLICT    342    343       MISSING (IN REF. 2).
SQ   SEQUENCE   511 AA;  57222 MW;  DE6D6C0533C7302F CRC64;

  Query Match             9.5%;  Score 494.5;  DB 1;  Length 511;
  Best Local Similarity   40.2%;  Pred. No. 9e-24;
  Matches  106;  Conservative   54;  Mismatches   95;  Indels    9;  Gaps    3;

Qy         51 HYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDN 110
              :|  :||:|:|:||  | |  :  :     ||:   || : : : :  | |:|| ::| |
Db          3 NYTVLRVIGQGSFGRALLVLQESSNQTFAMKEI---RLLKSDTQTSRKEAVLLAKMKHPN 59

Qy        111 IIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKA 170
              |:|:   |     | | :|||:||:|  :| :||  || |: :: :  ||   |: |||
Db         60 IVAFKESFEAEGYLYIVMEYCDGGDLMQRIKQQKGNLFPEDTILNWFIQICLGVNHIHKR 119

Qy        171 GILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNF 230
              :|||||: |:|||    :||||:| |: |:|  : | | |||||: ||: : : ||
Db        120 RVLHRDIKSKNVFLTHNGKVKLGDFGSARLLSSPMAFACTYVGTPYYVPPEIWENLPYNN 179

Qy        231 KSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLD 290
              |||||::||:::||  ||  | | :  || :|| ||   :    : || :|  :|   |
Db        180 KSDIWSLGCILYELCALKHPFQANSWKNLILKICQG--PIHPLPALYSCKLQGLVKQMLK 237

Qy        291 QDPEQRPTADELLDR----PLLRK 310
              ::|  ||:|  || |    ||: |
Db        238 RNPSHRPSATTLLCRGSLAPLVPK 261


RESULT 7
NEK2_MOUSE
ID   NEK2_MOUSE     STANDARD;      PRT;   443 AA.
AC   O35942; O35959;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase NEK2 (EC 2.7.1.37) (NimA-related
DE   protein kinase 2).
GN   NEK2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RC   STRAIN=Swiss Webster; TISSUE=Testis;
RX   MEDLINE=97330684; PubMed=9187143;
RA   Rhee K., Wolgemuth D.J.;
RT   "The NIMA-related kinase 2, Nek2, is expressed in specific stages of
RT   the meiotic cell cycle and associates with meiotic chromosomes.";
RL   Development 124:2167-2177(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98243037; PubMed=9583679;
RA   Arama E., Yanai A., Kilfin G., Motro B.;
RT   "Murine NIMA-related kinases are expressed in patterns suggesting
RT   distinct functions in gametogenesis and a role in the nervous
RT   system.";
RL   Oncogene 16:1813-1823(1998).
RN   [3]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Brain;
RX   MEDLINE=98096235; PubMed=9434622;
RA   Tanaka K., Parvinen M., Nigg E.A.;
RT   "The in vivo expression pattern of mouse Nek2, a NIMA-related kinase,
RT   indicates a role in both mitosis and meiosis.";
RL   Exp. Cell Res. 237:264-274(1997).
CC   -!- FUNCTION: PROTEIN KINASE THAT IS INVOLVED IN MITOTIC REGULATION.
CC       MAY HAVE A ROLE AT THE G2-M TRANSITION. MAY ALSO PLAY A ROLE IN
CC       MEIOSIS.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: Interacts with TERF1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANTLY EXPRESSED IN TESTIS. LOW
CC       LEVELS FOUND IN MID-GESTATION EMBRYO, OVARY, PLACENTA, INTESTINE,
CC       THYMUS AND SKIN. WITHIN THE TESTIS, EXRESSION RESTRICTED TO GERM
CC       CELLS WITH HIGHEST LEVELS DETECTED IN SPERMATOCYTES AT PACHYTENE
CC       AND DIPLOTENE STAGES. ALSO EXPRESSED IN MEIOTIC PACHYTENE OOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U95610; AAB67973.1; -.
DR   EMBL; AF013166; AAC35393.1; -.
DR   EMBL; AF007247; AAB70470.1; -.
DR   HSSP; P24941; 1CKP.
DR   MGD; MGI:109359; Nek2.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division; Meiosis.
FT   DOMAIN        8    271       PROTEIN KINASE.
FT   NP_BIND      14     22       ATP (BY SIMILARITY).
FT   BINDING      37     37       ATP (BY SIMILARITY).
FT   ACT_SITE    141    141       BY SIMILARITY.
FT   MOD_RES     179    179       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   443 AA;  51307 MW;  DE09565C378307E1 CRC64;

  Query Match             9.5%;  Score 493.5;  DB 1;  Length 443;
  Best Local Similarity   38.0%;  Pred. No. 8.5e-24;
  Matches  115;  Conservative   58;  Mismatches  109;  Indels   21;  Gaps    8;

Qy         46 EQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILAL 105
              :.| || |     | |::|     ||  |  ::||||:|   ::| |::  ::|: :|
Db          7 DYEVLHSI-----GTGSYGRCQKIRRKSDGKILVWKELDYGSMTEVEKQMLVSEVNLLRE 61

Qy        106 LQHDNIIAYYNHFMD--NTTLLIELEYCNGGNLYDKILR-QKDKLF-EEEMVVWYLFQIV 161
              |:| ||: ||:  :|  |||| | :||| ||:|   | :  ||: : ||| |:  : |:
Db         62 LKHPNIVRYYDRIIDRTNTTLYIVMEYCEGGDLASVISKGTKDRQYLEEEFVLRVMTQLT 121

Qy        162 SAVSCIHKAG-----ILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPY 216
               |:   |:       :||||:|  |:||   : :||||:|||: || : | |:| |||||
Db        122 LALKECHRRSDGGHTVLHRDLKPANVFLDSKHNVKLGDFGLARILNHDTSFAKTFVGTPY 181

Qy        217 YMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQ 276
              |||||    : || |||||:: |:::||  |   | | |   |  || :| |   :   :
Db        182 YMSPEQMSCLSYNEKSDIWSLACLLYELCALMPPFTAFNQKELAGKIREG-RFRRI-PYR 239

Qy        277 YSLELIQMVHSCLDQDPEQRPTADELLDRPLL-----RKRRREMEEKVTLLNAPTKRPRS 331
              ||  |  ::   |      ||: :|:|: ||:      ::|| :| :      |:| | |
Db        240 YSDGLNDLITRMLFLKDYHRPSVEEILESPLIADMVAEEQRRNLERRGRRSGEPSKLPDS 299

Qy        332 STV 334
              | |
Db        300 SPV 302


RESULT 8
NIMA_EMENI
ID   NIMA_EMENI     STANDARD;      PRT;   699 AA.
AC   P11837;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2-specific protein kinase NIMA (EC 2.7.1.-) (Never in mitosis).
GN   NIMA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88194523; PubMed=3359487;
RA   Osmani S.A., Pu R.T., Morris N.R.;
RT   "Mitotic induction and maintenance by overexpression of a G2-specific
RT   gene that encodes a potential protein kinase.";
RL   Cell 53:237-244(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196762; PubMed=7889945;
RA   Pu R.T., Osmani S.A.;
RT   "Mitotic destruction of the cell cycle regulated NIMA protein kinase
RT   of Aspergillus nidulans is required for mitotic exit.";
RL   EMBO J. 14:995-1003(1995).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE=95355415; PubMed=7629122;
RA   Pu R.T., Xu G., Wu L., Vierula J., O'Donnell K., Ye X.S.,
RA   Osmani S.A.;
RT   "Isolation of a functional homolog of the cell cycle-specific NIMA
RT   protein kinase of Aspergillus nidulans and functional analysis of
RT   conserved residues.";
RL   J. Biol. Chem. 270:18110-18116(1995).
CC   -!- FUNCTION: PROTEIN KINASE THAT PLAYS AN IMPORTANT ROLE IN MITOTIC
CC       REGULATION. SEEMS TO BE PHOSPHORYLATED AND THEREBY ACTIVATED BY
CC       CDC2/CYCLIN B DURING MITOTIC ACTIVATION. IT IS ALSO REQUIRED FOR
CC       SPINDLE FORMATION AND FOR NUCLEAR ENVELOPE BREAKDOWN.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES WHEN CELLS ARE ARRESTED IN G2
CC       AND IS DEGRADED AS CELLS TRAVERSE MITOSIS.
CC   -!- PTM: RECESSIVE MUTATIONS OF NIMA CAUSE A SPECIFIC CELL CYCLE BLOCK
CC       IN G2 AT RESTRICTIVE TEMPERATURE DUE TO LACK OF PHOSPHORYLATION
CC       THAT NORMALLY ACTIVATES THE G2-KINASE.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M20249; AAA33316.1; -.
DR   PIR; A43734; A43734.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division.
FT   DOMAIN       11    295       PROTEIN KINASE.
FT   NP_BIND      17     25       ATP (BY SIMILARITY).
FT   BINDING      40     40       ATP (BY SIMILARITY).
FT   ACT_SITE    166    166       BY SIMILARITY.
FT   MOD_RES     199    199       PHOSPHORYLATION (AUTO-).
FT   MUTAGEN      38     38       C->S,A: NO LOSS OF ACTIVITY.
FT   MUTAGEN      41     41       E->G: IN NIMA7; BLOCK IN G2.
FT   MUTAGEN      91     91       Y->N: IN NIMA5; BLOCK IN G2.
FT   MUTAGEN     199    199       T->A: GREAT DECREASE IN ACTIVITY.
FT   MUTAGEN     304    304       L->P: IN NIMA1; BLOCK IN G2.
SQ   SEQUENCE   699 AA;  78912 MW;  50FDC86E249ABC99 CRC64;

  Query Match             9.2%;  Score 476;  DB 1;  Length 699;
  Best Local Similarity   35.6%;  Pred. No. 2e-22;
  Matches  113;  Conservative   59;  Mismatches  109;  Indels   36;  Gaps    6;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
```

```
Db        11 YEVLEKIGCGSFGIIRKVKRKSDGFILCRKEINYIKMSTKEREQLTAEFNILSSLRHPNI 70

Qy       112 IAYYN--HFMDNTTLLIELEYCNGGNL--YDKILRQKDKLFEEEMVVWYLFQIVSAVSCI 167

Db        71 VAYYHREHLKASQDLYLYMEYCGGGDLSMVIKNLKRTNKYAEEDFVWRILSQLVTALYRC 130

Qy       168 H--------------------------KAGILHRDIKTLNIFLTKANLIKLGDYGLAKK 200

Db       131 HYGTDPAEVGSNLLGPAPKPSGLKGKQAQMTILHRDLKPENIFLGSDNTVKLGDFGLSKL 190

Qy       201 LNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLC 260

Db       191 MHS-HDFASTYVGTPFYMSPEICAAEKYTLRSDIWAVGCIMYELCQREPPFNARTHIQLV 249

Qy       261 VKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVT 320

Db       250 QKIREGKFAPLPDF--YSSELKNVIASCLRVNPDHRPDTATLINTPVIRLMRREVE--LN 305

Qy       321 LLNAPTKRPRSSTVTEA 337

Db       306 NLSRAARKREEATMQKA 322


RESULT 9
NRKA_TRYBB
ID   NRKA_TRYBB     STANDARD;      PRT;   431 AA.
AC   Q08942;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative serine/threonine-protein kinase A (EC 2.7.1.37).
GN   NRKA.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EATRO 164 / Isolate ISTAR1;
RX   MEDLINE=93295429; PubMed=8515773;
RA   Gale M.J. Jr., Parsons M.;
RT   "A Trypanosoma brucei gene family encoding protein kinases with
RT   catalytic domains structurally related to Nek1 and NIMA.";
RL   Mol. Biochem. Parasitol. 59:111-122(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L03778; AAB59252.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00169; PH; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN       20    279       PROTEIN KINASE.
FT   DOMAIN      331    429       PH.
FT   NP_BIND      26     34       ATP (BY SIMILARITY).
FT   BINDING      49     49       ATP (BY SIMILARITY).
FT   ACT_SITE    147    147       BY SIMILARITY.
FT   VARIANT     192    192       A -> P (IN STRAIN TREU66).
FT   VARIANT     199    199       K -> L (IN STRAIN TREU66).
SQ   SEQUENCE   431 AA;  47915 MW;  2E68B70275884224 CRC64;

  Query Match             8.9%;  Score 461.5;  DB 1;  Length 431;
  Best Local Similarity   29.9%;  Pred. No. 8.3e-22;
  Matches  127;  Conservative   77;  Mismatches  152;  Indels   69;  Gaps   13;

Qy        37 GPRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDA 96

Db        10 GTDGSGGRCK-----YLNKGIVGLGSYGEAYVAESVEDGSLCVAKVMDLSKMSQRDKRYA 64

Qy        97 LNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKI-LRQKD--KLFEEEMV 153

Db        65 QSEIKCLANCNHPNIIRYIEDHEENDRLLIVMEFADSGNLDEQIKLRGSGDARYFQEHEA 124

Qy       154 VWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYS--MAETL 211

Db       125 LFLFLQLCLALDYIHSHKMLHRDIKSANVLLTSTGLVKLGDFGFSHQYEDTVSGVVASTF 184

Qy       212 VGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME 271

Db       185 CGTPYYLAPELWNNKRYNKKADVWSLGVLLYEIMGMKKPFSASNLKGLMSKVLAGTYAPL 244

Qy       272 VDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREMEEKVTLLNAPTKRPR- 330

Db       245 PDS--FSSEFKRVVDGILVADPNDRPSVREIFQIPYINKGLK------LFVQALKKNERI 296

Qy       331 SSTVTEAPIAVVTSRTSEVYV----------------------------WGGGKS-TPQ 360

Db       297 SDSVKE----VLVTQVSEILSSEVSPDAHRFLVSQINYDVTHRGHVNKLGGGNGKSWKPR 352

Qy       361 KLDVIK--------------SGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQ- 405

Db       353 FLQIVRGQLILTDDEEGNNPKGLNLEQV-QGACPVPHSTAKRDFVFALNTVGGKGMWFQA 411

Qy       406 LGHGD 410

Db       412 VSHGD 416


RESULT 10
NIM1_NEUCR
ID   NIM1_NEUCR     STANDARD;      PRT;   779 AA.
AC   P48479;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2-specific protein kinase nim-1 (EC 2.7.1.-).
GN   NIM-1.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95355415; PubMed=7629122;
RA   Pu R.T., Xu G., Wu L., Vierula J., O'Donnell K., Ye X.S.,
RA   Osmani S.A.;
RT   "Isolation of a functional homolog of the cell cycle-specific NIMA
RT   protein kinase of Aspergillus nidulans and functional analysis of
RT   conserved residues.";
RL   J. Biol. Chem. 270:18110-18116(1995).
CC   -!- FUNCTION: PROTEIN KINASE THAT PLAYS AN IMPORTANT ROLE IN MITOTIC
CC       REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES WHEN CELLS ARE ARRESTED IN G2;
CC       DEGRADED AS CELLS TRAVERSE MITOSIS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       NIMA SUBFAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L42573; AAA80145.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW   Nuclear protein; Phosphorylation; Cell cycle; Cell division.
FT   DOMAIN        7    290       PROTEIN KINASE.
FT   NP_BIND      13     21       ATP (BY SIMILARITY).
FT   BINDING      36     36       ATP (BY SIMILARITY).
FT   ACT_SITE    161    161       BY SIMILARITY.
FT   MOD_RES     194    194       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   DOMAIN      751    759       POLY-SER.
SQ   SEQUENCE   779 AA;  86079 MW;  7C174925A25EB9B9 CRC64;

  Query Match             8.9%;  Score 461.5;  DB 1;  Length 779;
  Best Local Similarity   36.1%;  Pred. No. 1.9e-21;
  Matches  107;  Conservative   52;  Mismatches  102;  Indels   35;  Gaps    8;

Qy         52 YIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNI 111
              |  :  :| |:||     ||  |  ::  ||:   ::|:|||     |  ||: |:| ||
Db          7 YELLEKIGHGSFGIIRKVRRKADGMILCRKEISYLKMSQKEREQLHAEFSILSTLRHPNI 66

Qy        112 IAYYN--HFMDNTTLLIELEYCNGGNLYDKILRQ--KDKLFEEEMVVWYLF-QIVSAVSC 166
              : ||:  |      | : :|||  |:|  :::|   |:  : ||  || :| |:|:|:
Db         67 VGYYHREHLKATQDLHLYMEYCGNGDL-GRVIRNLIKNNQYAEESFVWSIFSQLVTALYR 125

Qy        167 IH------KAG--------------------ILHRDIKTLNIFLTKANLIKLGDYGLAKK 200
              |      : |                    |||||:|  |:|| : | :||||:||:|
Db        126 CHYGVDPPEVGKTVLGLGSTARPKPPSGGMTILHRDLKPENVFLGEDNSVKLGDFGLSKV 185

Qy        201 LNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLC 260
              : | :  | | ||||:|||||:|   ||  |||||::||:|:||   :  |:|      |
Db        186 MQS-HDFASTYVGTPFYMSPEICAAEKYTLKSDIWSLGCIIYELCAREPPFNAKTHYQLV 244

Qy        261 VKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREME 316
               || :|   :    | || ||   :  ||  :|::||    ||: |::|  |:| |
Db        245 QKIKEG--KIAPLPSVYSGELFATIKDCLRVNPDRRPDTATLLNLPIVRLMRKEKE 298


RESULT 11
NRKB_TRYBB
ID   NRKB_TRYBB     STANDARD;      PRT;   431 AA.
AC   Q03428;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative serine/threonine-protein kinase B (EC 2.7.1.37).
GN   NRKB.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate TREU66;
RX   MEDLINE=93295429; PubMed=8515773;
RA   Gale M.J. Jr., Parsons M.;
RT   "A Trypanosoma brucei gene family encoding protein kinases with
RT   catalytic domains structurally related to Nek1 and NIMA.";
RL   Mol. Biochem. Parasitol. 59:111-122(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L03777; AAB59253.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00169; PH; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN       20    279       PROTEIN KINASE.
FT   DOMAIN      331    429       PH.
FT   NP_BIND      26     34       ATP (BY SIMILARITY).
FT   BINDING      49     49       ATP (BY SIMILARITY).
FT   ACT_SITE    147    147       BY SIMILARITY.
SQ   SEQUENCE   431 AA;  48172 MW;  F6B4B3EF9E99FF74 CRC64;

  Query Match             8.8%;  Score 454.5;  DB 1;  Length 431;
  Best Local Similarity   29.4%;  Pred. No. 2.3e-21;
  Matches  122;  Conservative   76;  Mismatches  168;  Indels   49;  Gaps   10;

Qy         37 GPRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDA 96
              |    ||  :     |:   ::| |::||  :  | || || | | :||:::| :::| |
Db         10 GTDGSGGRCK-----YLNKGIVGLGSYGEGYVAERVEDGSLCVAKVMDLSKMSRRDKRYA 64

Qy         97 LNEIVILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKIL---RQKDKLFEEEMV 153
              :||      | ||| |     :|  ||| :|: : ||| ::|       : |:|
Db         65 QSEIKYPTNCNHPNIIRYIEDHEENDRLLIVMEFADSGNLDEQIKPWGTGDARYFQEHEA 124

Qy        154 VWYLFQIVSAVSCIHKAGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSEYS--MAETL 211
              ::   |:  |:  ||   :|||||: |: ||   |:||||:| : :     |  :| |
Db        125 LFLFLQLCLALDYIHSHKMLHRDIKSANVLLTSTGLVKLGDFGFSHQYEDTVSGVVASTF 184

Qy        212 VGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME 271
              |||||::|||   ::|| |:|:|:|:| :::|:: :|: | |:|   |  |:: |  |
Db        185 CGTPYYLAPELWNNLRYNKKADVWSLGVLLYEIMGMKKPFSASNLKGLMSKVLAGTYAPL 244

Qy        272 VDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRP-----------LLRKRRREMEEKVT 320
              ||  :| |  ::|   |  ||  ||:  |    |             |:|  | ::
Db        245 PDS--FSSEFKRVVDGILVADPNDRPSVRENFQIPYINKGLKLFVQALKKNERILDSVKE 302

Qy        321 LLNAPTKRPRSSTVT-------EAPIAVVTSRTSEVYVWGGGKS---TPQKLDVIK---- 366
              :| :      || |:       |: |    :    |   |||      |: | :::
Db        303 VLVSQVSEILSSEVSPDAHRFLESQINYDVTHRGHVNKLGGGNGKSWKPRFLQIVRGQLI 362

Qy        367 ----------SGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQ-LGHGD 410
                       | :  ||  |       | :::    :|  ||  :  | : |||
Db        363 LTDDEEGNNPKGLNLEQV-QGACPVPYSTAKRDFVFALNTVGGEGMWFQAVSHGD 416


RESULT 12
KIN3_YEAST
ID   KIN3_YEAST     STANDARD;      PRT;   435 AA.
AC   P22209;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine-protein kinase KIN3 (EC 2.7.1.-).
```

```
GN   KIN3 OR NPK1 OR YAR018C OR FUN52.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92354926; PubMed=1644305;
RA   Barton A.B., Davies C.J., Hutchison C.A. III, Kaback D.B.;
RT   "Cloning of chromosome I DNA from Saccharomyces cerevisiae: analysis
RT   of the FUN52 gene, whose product has homology to protein kinases.";
RL   Gene 117:137-140(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92357012; PubMed=1495480;
RA   Schweitzer B., Philippsen P.;
RT   "NPK1, a nonessential protein kinase gene in Saccharomyces cerevisiae
RT   with similarity to Aspergillus nidulans nimA.";
RL   Mol. Gen. Genet. 234:164-167(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95028152; PubMed=7941740;
RA   Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
RA   Delaney S., Ouellette B.F.F., Barton A.B., Kaback D.B., Bussey H.;
RT   "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
RT   the 42 kbp SPO7-CENI-CDC15 region.";
RL   Yeast 10:535-541(1994).
RN   [4]
RP   SEQUENCE OF 1-369 FROM N.A.
RC   STRAIN=SB303;
RX   MEDLINE=90337351; PubMed=2199332;
RA   Jones D.G.L., Rosamond J.;
RT   "Isolation of a novel protein kinase-encoding gene from yeast by
RT   oligodeoxyribonucleotide probing.";
RL   Gene 90:87-92(1990).
CC   -!- FUNCTION: THIS PROTEIN IS PROBABLY A SERINE/THREONINE PROTEIN
CC       KINASE.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STRONG, TO A.NIDULANS BIMA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67445; AAA34399.2; -.
DR   EMBL; X60549; CAA43042.1; -.
DR   EMBL; L22015; AAC04964.1; -.
DR   EMBL; M55416; AAB04029.1; -.
DR   PIR; S11185; S11185.
DR   PIR; S20123; S20123.
DR   PIR; S23580; S23580.
DR   PIR; S43451; S43451.
DR   SGD; S0000071; KIN3.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 2.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN       25    343       PROTEIN KINASE.
FT   NP_BIND      31     39       ATP (BY SIMILARITY).
FT   BINDING      54     54       ATP (BY SIMILARITY).
FT   ACT_SITE    174    174       BY SIMILARITY.
FT   CONFLICT     92     92       F -> L (IN REF. 4).
FT   CONFLICT     98     98       V -> L (IN REF. 4).
FT   CONFLICT    121    121       E -> Q (IN REF. 4).
FT   CONFLICT    130    130       V -> I (IN REF. 4).
FT   CONFLICT    235    238       AKSL -> SQIS (IN REF. 4).
FT   CONFLICT    358    358       K -> R (IN REF. 4).
SQ   SEQUENCE   435 AA;  51203 MW;  7CB49CD6A93CA019 CRC64;

  Query Match             7.7%;  Score 397;  DB 1;  Length 435;
  Best Local Similarity   30.2%;  Pred. No. 9.4e-18;
  Matches  106;  Conservative   53;  Mismatches  114;  Indels   78;  Gaps    8;

Qy         32 PSASQG--PRAGGGAAEQEELHYIPIRVLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLS 89
              |   ||  ||:           |  :  :|||:||           |:| |::    ::
Db         13 PQQQQGHPPRS----------EYQVLEEIGRGSFGSVRKVIHIPTKKLLVRKDIKYGHMN 62

Qy         90 EKERRDALNEIVILALLQHDNIIAYYNHFMD--NTTLLIELEYCNGGNLYDKI--LRQKD 145
              |||:  : |  ||: |:|:||: :||   |     | : :|||: |:|   | . :|:
Db         63 SKERQQLIAECSILSQLKHENIVEFYNWDFDEQKEVLYLYMEYCSRGDLSQMIKHYKQEH 122

Qy        146 KLFEEEMVVWYLFQIVSAVSCIH--------------------KAGILHRDIKTLNIFL 184
              |   |::|   | |:::|:   |                     |  ::|||:|  ||||
Db        123 KYIPEKIVWGILAQLLTALYKCHYGVELPTLTTIYDRMKPPVKGKNIVIHRDLKPGNIFL 182

Qy        185 T-------------------------KAN-------------LIKLGDYGLAKKLNSEY 205
              :                         : |             ::||||:|||| | :
Db        183 SYDDSDYNINEQVDGHEEVNSNYYRDHRVNSGKRGSPMDYSQVVVKLGDFGLAKSLETSI 242

Qy        206 SMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQ 265
              | | |||||||||:     |:  ||||::||||||: :|   | | | | |  ||
Db        243 QFATTYVGTPYYMSPEVLMDQPYSPLSDIWSLGCVIFEMCSLHPPFQAKNYLELQTKIKN 302

Qy        266 GIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTADELLDRPLLRKRRREME 316
              |    :     ||  |  ::|| :| :   ||:  |||    :|  |: ::
Db        303 G--KCDTVPEYYSRGLNAIIHSMIDVNLRTRPSTFELLQDIQIRTARKSLQ 351


RESULT 13
YLK3_CAEEL
ID   YLK3_CAEEL     STANDARD;      PRT;  1576 AA.
AC   P41951; Q95QP7;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative serine/threonine-protein kinase D1044.3 in chromosome III
DE   (EC 2.7.1.-).
GN   D1044.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Pauley A., Waterston R.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Waterston R.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a (shown here) and b; are
CC       produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STRONG, TO ZC84.1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00065; AAK68286.1; -.
DR   EMBL; U00065; AAL27237.1; -.
DR   HSSP; Q63450; 1A06.
```

```
DR   WormPep; D1044.3a; CE27894.
DR   WormPep; D1044.3b; CE29743.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR002899; WR1/EB.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01683; EB; 12.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00289; WR1; 12.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW   ATP-binding; Alternative splicing.
FT   DOMAIN      431    703       PROTEIN KINASE.
FT   NP_BIND     437    445       ATP (BY SIMILARITY).
FT   BINDING     461    461       ATP (BY SIMILARITY).
FT   ACT_SITE    569    569       BY SIMILARITY.
FT   VARSPLIC      1    904       MISSING (IN ISOFORM B).
FT   VARSPLIC    905   1012       SPEEIQMNLGFSDFRPVLVQSHINGYQKDQKVTKITYEQLS
FT                                ACIECLIAENPAAKHVPHRTRAVVILRRDLLVLGQYVNMLV
FT                                PTITTYVVIRQIHVSLAAILVYTEYE -> MEETCSESPKE
FT                                SNIISFIWHKLLKRVPPPIMICLFFLLLQIFVISVVSQCPP
FT                                GLTPLFSNSNFNQPLTCTPQDACSCYSSSGSSRFGTICQYA
FT                                STYNNYICCYSTNTQ (IN ISOFORM B).
SQ   SEQUENCE   1576 AA;  174677 MW;  3A11EE573E813498 CRC64;

  Query Match             7.4%;  Score 381.5;  DB 1;  Length 1576;
  Best Local Similarity   23.6%;  Pred. No. 5.1e-16;
  Matches  162;  Conservative  117;  Mismatches  237;  Indels  169;  Gaps   29;

Qy         39 RAGGGAAEQEELHYIPIRVLGRGAFG-EATLYRRTEDDS-----LVVWKEVDLTRLSEKE 92
              :| |   | | |       || ||||   |: :: :  |     |:  ||: :| |:::|
Db        423 QAVGEVGEYELLDQ-----LGAGAFGCVYTVRKKAQSHSENPAKLLALKEIFMTNLNDRE 477

Qy         93 R----RDALNEI-VILALLQHDNIIAYYNHFMDNTTLLIELEYCNGGNLYDKILRQKDKL 147
                  | ::|: :|   |:| ||: |   |::|  | | ::   | :| | |:  |:|
Db        478 SDKSFGDMISEVKIIKQQLRHPNIVRYRRIFVENHRLYIVMDLIQGCSLRDLIITMKEKK 537

Qy        148 --FEEEMVVWYLFQIVSAVSCIHK-AGILHRDIKTLNIFLTKANLIKLGDYGLAKKLNSE 204
               |||: :   : |:: |:  :||   |:|||:|  || :|    : : |:||||:   |
Db        538 GNFEEKKIWAMVVQMMLALRYLHKEKQIVHRDLKPNNIMMTTDERVVITDFGLAKQKGPE 597

Qy        205 YSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCVIFELLTLKRTFDATNPLNLCVKIV 264
              |   ::  ||  |  ||: | : |  |:|||: || |:|:  |:  | :|| | | ::||
Db        598 Y--LKSAAGTIIYSCPEIVQNLPYGEKADIWSFGCCIYEMCQLQPVFHSTNMLTLAMQIV 655

Qy        265 QGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRP-----------TADELLDRPLLRKRRR 313
              :     :  :| :|  :: |||  ||  ||              | ||  : |::
Db        656 EA--KYDPLNEMWSDDLRFLITSCLAPDPSARPDILKVSGMCGVRLLEYLD-DVARQQAS 712

Qy        314 -------------EMEEKVTLLNAPT---KRPRSSTVT---------------------- 335
                          :::|  : ||: |   |||  |: |
Db        713 TSDMTASQSSYNIKIDESPSSLNSSTSSYKRPGRSSKTSGSGKLPPINPAPRRNHSMSAG 772

Qy        336 ----------------------EAPIAVVTSR----TSEVYVWGGGKSTPQKLDVIKSG 368
                                   || |: : |     |  :     ||  |:      |
Db        773 ETPRPSSIVCLPRITDKYSVMFPSAPSAIPSRRRVQTCSTEHPARSSSSTELKVSKQSDG 832

Qy        369 CSA-----RQV------CAGNTHFAVVTVEKELYTWVNMQGGTKLHGQ----------L 406
              :      ||:           |  :|  :||  :       | :: |           |
Db        833 LTVSSNVLRQIQDPVLTILNQIHRILVVTDKETIS-------TSMNHQRRLVEMFRKNLL 885

Qy        407 GHGDKASYRQPKHVEKLQGKAIHQVSCG---DDFTVCVTDEGQLYAFGSDYYGCMGVDKV 463
              |  : |  :   |: ||  ::  ::       ||   |  :  :  :  |      | |:
Db        886 GRENDA-VQMKTHLRKLAAESPEEIQMNLGFSDFRP-VLVQSHINGYQKD----QKVTKI 939

Qy        464 AGPEVLEPMQLNFFLSNP-VEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYT 522
              ::   ::      ||  : |      ||:| |:  |     |:|  : : :  |
Db        940 TYEQLSACIEC-LIAENPAAKHVPHRTRAVVILRRDLLV----LGQYVNMLVPTITTYVV 994

Qy        523 PQKVDVPKALIIV--AVQCGCDGTFLLTQSGKVLACGL--------------------N 559
                                                                        |
Db        995 IRQIHVSLAAILVYTEYECGSNSSPQVSASGQVVTCSTNTQCASGYTCNNGACCPNTNSN 1054

Qy        560 EFNKLGLNQCMSG--IINHEAYHEV 582
                :  | | |::|  ::| : |: |
Db       1055 TCSSNGNNGCLAGQTMVNGQCYNSV 1079


RESULT 14
NRK1_YEAST
ID   NRK1_YEAST     STANDARD;      PRT;  1080 AA.
AC   P38692;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine-protein kinase NRK1 (EC 2.7.1.37) (N-rich kinase 1).
GN   NRK1 OR KIC1 OR YHR102W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC-5;
RA   Fukami Y.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D29980; BAA06250.1; -.
DR   EMBL; U00059; AAB68860.1; -.
DR   PIR; S48944; S48944.
DR   SGD; S0001144; KIC1.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN      23    276       PROTEIN KINASE.
FT   NP_BIND     29     37       ATP (BY SIMILARITY).
FT   BINDING     52     52       ATP (BY SIMILARITY).
FT   ACT_SITE   144    144       BY SIMILARITY.
SQ   SEQUENCE   1080 AA;  117061 MW;  9989EAF315EE0B94 CRC64;

  Query Match             7.3%;  Score 377;  DB 1;  Length 1080;
  Best Local Similarity   21.1%;  Pred. No. 5.8e-16;
  Matches  206;  Conservative  138;  Mismatches  390;  Indels  240;  Gaps   36;
```

```
Qy         57 VLGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILA-LLQHDNIIAYY 115
              |:||| ||      :   :   | ::|   |: |  |   ||  || | |  ||  ||
Db         28 VIGRGKFGVVYKGYNVKTGRVYAIKVLNLDSDSD-EVEDVQREIQFLASLKQISNITRYY 86

Qy        116 NHFMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHR 175
              :: :|:| | :|:| ||:|  : | :  |: :|: :   : ::: |: |||| ::||
Db         87 GSYLKDTSLWIIMEHCAGGSL--RSLLRPGKI-DEKYIGVIMRELLVALKCIHKDNVIHR 143

Qy        176 DIKTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPE-LCQGVKYNFKSDI 234
              |||  |: :|    :|| |:|:| ::|      :|: ||||:|:|| : :|| |: | ||
Db        144 DIKAANVLITNEGNVKLCDFGVAAQVNQTSLRRQTMAGTPYWMAPEVIMEGVYYDTKVDI 203

Qy        235 WAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAME---------VDSSQYSLELIQMV 285
              |::|,  :|: |        ||   |   |: :|||:          ::   ||  | : :
Db        204 WSLGITTYEIAT-------GNP-PYC--DVEALRAMQLIIKSKPPRLEDRSYSTSLKEFI 253

Qy        286 HSCLDQDPEQRPTADELLDRPLLRKRR-------REMEEKVTLL---------------- 322
              |||:||::| :|||:||    :|  :       :|:  :  |
Db        254 ALCLDEDPKERLSADDLLKSKFIRAHKATPTSILKELISRYLLFRDKNKNKYKIEGSIPE 313

Qy        323 NAPTK-----RPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQKLD---------VIKSG 368
              | |:|     :|  :   :     : | :|:      |     | |          :|::
Db        314 NEPSKPSEAPKPSQNGGGDEAQKSIASNDNEIKRVNEG-DVEMKWDFDSLSSSDYIIENN 372

Qy        369 CSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAI 428
              :   :  |  :|  | : :|: :          . : |: ||        :
Db        373 INLDALAEDNNEWA--TAQHDLFNY------------AYPDEDSYYFDPTSHNTRPFVY 417

Qy        429 HQVSCGDDFTVCVTDEGQLYA------FGSDYYGCM--GVDKVAGPEVLEPMQLNFFLSN 480
              : |  :   :    | |        | |   |  |    :|     ||  :  | :
Db        418 QGTTIGKGYPGTIAQNSTLNAPVTNNYTNSKYPSKMVAGTTNTSGTHTAGPMTSSKRLES 477

Qy        481 --PVEQVSCGDNHVVVLTRN----------KEVYSWGCGEYGRLGLDSEEDYYTP----- 523
              | : :   ::: ::   |          | : |   |       :|  | : |
Db        478 KAPKQLLELFEDNEIITAENDVNTEAPKISKSISSLNAG-------NSSRDDFIPSISNE 530

Qy        524 ---------QKVDVPKALIIVAVQCGCDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGI- 573
                    :  :|          |   :||  |     :|||        ||:
Db        531 VNGNINNNKMRPHLPPL---------SSGNNYYSQSTPALPLLQTKFNKTSKGPPTSGLT 581

Qy        574 -----INHEAYHEVPYTTSFTLAKQLSFYKIRTIAPGKTHTAAIDERGRLLTFGCNKCGQ 628
                  |  |   |:| :   | |         | |   | ||        : :
Db        582 TAPTSIEIEIPEELPNSALPTPASADPVLIPSTKARSSTVTAGTPSSSSSIQY------- 634

Qy        629 LGVGNYKKRLGIN----------LLGGPLGGKQVIRVSCGDEFTIAATDDNHIFAWGNGG 678
                |  :|| ::          :    ||                |   :: | :  |
Db        635 KSPSNVPRRLTVSNNRPEHCPSTITNQKLGS-------------AVASNSGISSTPNNS 680

Qy        679 NGRLAMTPTERPHGSD-----------------------ICTSWPRPIFG---SLHHVP 711
              |     | :|   ||                      :|  ||:||   | :  |
Db        681 NNYNNNTDSENSRGSSGSNTANSTQMGITNPGNVTKLSTHKASSPSRPLFGVGTSPNRKP 740

Qy        712 ------DLSCRGWHTIL--------------IVEKVLNS-------KTIRSNSSGLS 741
                   ::     || |              |: : |||         ||  |||
Db        741 AGSPTQNIGHNSTHTNLAPPPTMKPMANSKDNKDILLQPLNSIPSSSTLNTISGNSSNNL 800

Qy        742 IGTVFQSSSPGGGGGGGGEEEDSQQESETPDPSGGFRGTMEADRGMEGLISPTEAMG-N 800
              : : |:   |    |  :   : : | |:   |           |: |  |   |
Db        801 TSSNYFSNEKEGSRVNGDFKRNNPNLKLQMPLPTPVVRNK---------LLDPNTATSQN 851

Qy        801 SNGASSSCPGWLRKELENAEFIPMPDSPSPLS-AAFSESEKDTLPYEELQGLKVASEAPL 859
              :||   |     : :    |     | |:|    ||  |       | |:     :
Db        852 NNGMPGSAGISTNENINQFGFNTSSASNIPVSMTPISEKHIDF-------GGKIKRSQSI 904

Qy        860 EHKPQVEASSPRLN 873
               ::    ||  ||
Db        905 SNRKNSSASEHPLN 918


RESULT 15
ST10_HUMAN
ID   ST10_HUMAN     STANDARD;      PRT;   968 AA.
AC   O94804; Q9UIW4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine-protein kinase 10 (EC 2.7.1.37) (Lymphocyte-oriented
DE   kinase).
GN   STK10 OR LOK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99216434; PubMed=10199912;
RA   Kuramochi S., Matsuda Y., Okamoto M., Kitamura F., Yonekawa H.,
RA   Karasuyama H.;
RT   "Molecular cloning of the human gene STK10 encoding lymphocyte-
RT   oriented kinase, and comparative chromosomal mapping of the human,
RT   mouse, and rat homologues.";
RL   Immunogenetics 49:369-375(1999).
RN   [2]
RP   SEQUENCE OF 814-968 FROM N.A.
RC   TISSUE=Testis;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CAN ACT ON SUBSTRATES SUCH AS MYELIN BASIC PROTEIN AND
CC       HISTONE IIA ON SERINE AND THREONINE RESIDUES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN LYMPHOID ORGANS.
CC   -!- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STE20 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015718; BAA35073.1; -.
DR   EMBL; AL133081; CAB61400.1; -.
DR   HSSP; P24941; 1HCL.
DR   Genew; HGNC:11388; STK10.
DR   MIM; 603919; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Coiled coil.
FT   DOMAIN       36    294       PROTEIN KINASE.
FT   DOMAIN      573    947       COILED COIL (POTENTIAL).
FT   DOMAIN      750    884       GLN-RICH.
FT   NP_BIND      42     50       ATP (BY SIMILARITY).
FT   BINDING      65     65       ATP (BY SIMILARITY).
FT   ACT_SITE    157    157       BY SIMILARITY.
SQ   SEQUENCE   968 AA;  112134 MW;  15E245193ECC553D CRC64;

  Query Match             6.9%;  Score 357.5;  DB 1;  Length 968;
  Best Local Similarity   26.1%;  Pred. No. 8.4e-15;
  Matches  114;  Conservative   78;  Mismatches  166;  Indels   79;  Gaps   14;

Qy         58 LGRGAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNH 117
              || ||||:    :  |  :|   | ::    ||:|  | : || |||   |  |:
```

```
Db         42 LGDGAFGKVYKAKNKETGALAAAKVIETK--SEEELEDYIVEIEILATCDHPYIVKLLGA 99

Qy        118 FMDNTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDI 177

Db        100 YYHDGKLWIMIEFCPGGAV-DAIMLELDRGLTEPQIQVVCRQMLEALNFLHSKRIIHRDL 158

Qy        178 KTLNIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPE--LCQGVK---YNFKS 232

Db        159 KAGNVLMTLEGDIRLADFGVSAKNLKTLQKRDSFIGTPYWMAPEVVMCETMKDTPYDYKA 218

Qy        233 DIWAVGCVIFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQD 292

Db        219 DIWSLGITLIEMAQIEPPHHELNPMRVLLKIAKSDPPTLLTPSKWSVEFRDFLKIALDKN 278

Qy        293 PEQRPTADELLDRPLL-----RKRRREM------------------EEKVTLLNAPTKR 328

Db        279 PETRPSAAQLLEHPFVSSITSNKALRELVAEAKAEVMEEIEDGRDEGEEEDAVDAASTLE 338

Qy        329 PRSSTVTE-APIAVVTSRTSEVYVWGGGKSTP----QKLDVIKSGCSARQ---------- 373

Db        339 NHTQNSSEVSPPSLNADKPLE-----ESPSTPLAPSQSQDSVNEPCSQPSGDRSLQTTSP 393

Qy        374 --VCAGNTHFAVVTV-----------------EKELYTWVNMQGGTKLHGQLGHGDKAS 413

Db        394 PVVAPGNENGLAVPVPLRKSRPVSMDARIQVAQEKQ----VAEQGG-DLSPAANRSQKAS 448

Qy        414 YRQPKH--VEKLQGKAI 428

Db        449 QSRPNSSALETLGGEKL 465
```

Search completed: July 1, 2003, 08:31:15
Job time : 21.9185 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 1, 2003, 08:30:24 ; Search time 15.0815 Seconds
(without alignments)
2266.116 Million cell updates/sec

Title: US-09-884-001-2
Perfect score: 4095
Sequence: 1 MSAPSEEEEYARLVMEAQPE.....QTRRGRAKAAPKTKPATPSL 824

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1330	32.5	782	1	BICD_DROME	P16568 drosophila
2	358.5	8.8	1938	1	MYS_AEQIR	P24733 aequipecten
3	352.5	8.6	1790	1	USO1_YEAST	P25386 saccharomyc
4	352	8.6	1972	1	MYHB_HUMAN	P35749 homo sapien
5	346.5	8.5	2116	1	MYS2_DICDI	P08799 dictyosteli
6	346	8.4	1976	1	MYHA_HUMAN	P35580 homo sapien
7	345	8.4	1976	1	MYHA_RAT	Q9jlt0 rattus norv
8	344.5	8.4	1972	1	MYHB_RABIT	P35748 oryctolagus
9	341	8.3	1976	1	MYHA_BOVIN	Q27991 bos taurus
10	338.5	8.3	1978	1	MYHB_CHICK	P10587 gallus gall
11	337.5	8.2	1961	1	MYH9_RAT	Q62812 rattus norv
12	335.5	8.2	1972	1	MYHB_MOUSE	O08638 mus musculu
13	335	8.2	1960	1	MYH9_HUMAN	P35579 homo sapien
14	331	8.1	1962	1	MYSA_DROME	P05661 drosophila
15	331	8.1	2017	1	MYSN_DROME	Q99323 drosophila
16	328.5	8.0	1959	1	MYH9_CHICK	P14105 gallus gall
17	326.5	8.0	1130	1	YL17_CAEEL	Q11102 caenorhabdi
18	326.5	8.0	4473	1	PLE1_CRIGR	Q9ji55 cricetulus
19	325	7.9	2663	1	CENE_HUMAN	Q02224 homo sapien
20	323.5	7.9	1102	1	MYSC_CHICK	P29616 gallus gall
21	323.5	7.9	1957	1	YD86_SCHPO	Q10411 schizosacch
22	323	7.9	1433	1	REST_CHICK	O42184 gallus gall
23	321.5	7.9	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
24	321.5	7.9	1935	1	MYH7_HUMAN	P12883 homo sapien
25	320	7.8	1940	1	MYH3_CHICK	P02565 gallus gall
26	319.5	7.8	1940	1	MYH3_HUMAN	P11055 homo sapien
27	318.5	7.8	1935	1	MYH7_PIG	P79293 sus scrofa
28	317.5	7.8	1934	1	MYH7_MESAU	P13540 mesocricetu
29	317.5	7.8	1935	1	MYH7_RAT	P02564 rattus norv
30	317.5	7.8	1940	1	MYH3_RAT	P12847 rattus norv
31	317	7.7	1084	1	MYSS_RABIT	P02562 oryctolagus
32	315	7.7	1938	1	MYSS_CHICK	P13538 gallus gall
33	314.5	7.7	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
34	314.5	7.7	4684	1	PLE1_HUMAN	Q15149 homo sapien
35	314	7.7	4687	1	PLE1_RAT	P30427 rattus norv
36	310.5	7.6	1939	1	MYH1_HUMAN	P12882 homo sapien
37	310.5	7.6	1939	1	MYH4_HUMAN	Q9y623 homo sapien
38	308.5	7.5	1325	1	G160_MOUSE	P55937 mus musculu
39	308	7.5	879	1	MYSP_ONCVO	Q02171 onchocerca
40	308	7.5	1941	1	MYH2_HUMAN	Q9ukx2 homo sapien
41	305.5	7.5	1938	1	MYH4_RABIT	Q28641 oryctolagus
42	304.5	7.4	1939	1	MYH6_HUMAN	P13533 homo sapien
43	304	7.4	1875	1	MLP1_YEAST	Q02455 saccharomyc
44	303	7.4	3210	1	CENF_HUMAN	P49454 homo sapien
45	301	7.4	1938	1	MYH6_MOUSE	Q02566 mus musculu

ALIGNMENTS

RESULT 1
BICD_DROME
ID BICD_DROME STANDARD; PRT; 782 AA.
AC P16568; Q9VJD5;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytoskeleton-like bicaudal D protein.
GN BICD OR CG6605.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90075232; PubMed=2590944;
RA Wharton R.P., Struhl G.;
RT "Structure of the Drosophila BicaudalD protein and its role in
RT localizing the the posterior determinant nanos.";
RL Cell 59:881-892(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=90152340; PubMed=2576013;
RA Suter B., Romberg L.M., Steward R.;
RT "Bicaudal-D, a Drosophila gene involved in developmental asymmetry:
RT localized transcript accumulation in ovaries and sequence similarity
RT to myosin heavy chain tail domains.";
RL Genes Dev. 3:1957-1968(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR DIFFERENTIATION. IT
CC       MAY PLAY A ROLE IN LOCALIZING OF NANOS (A MATERNAL DETERMINANT)
CC       ACTIVITY IN OOCYTES. BICD MUTATIONS CAUSE NANOS MISLOCALIZATION
CC       AND THUS BICAUDAL DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: OOGENESIS.
CC   -!- SIMILARITY: OF C-TERMINAL HALF TO VARIOUS MYOSIN HEAVY CHAINS AND
CC       SOME INTERMEDIATE FILAMENT PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31684; AAA28393.1; -.
DR   EMBL; X51652; CAA35964.1; -.
DR   EMBL; AE003655; AAF53616.1; -.
DR   PIR; A33636; A33636.
DR   FlyBase; FBgn0000183; BicD.
KW   Coiled coil; Developmental protein.
FT   DOMAIN        3    263       COILED COIL (POTENTIAL).
FT   DOMAIN      319    477       COILED COIL (POTENTIAL).
FT   DOMAIN      601    746       COILED COIL (POTENTIAL).
FT   CONFLICT    296    296       A -> S (IN REF. 1).
FT   CONFLICT    318    318       L -> P (IN REF. 1).
FT   CONFLICT    477    477       H -> R (IN REF. 1).
SQ   SEQUENCE   782 AA;  88953 MW;  5A71776171DF58E6 CRC64;

  Query Match          32.5%;  Score 1330;  DB 1;  Length 782;
  Best Local Similarity   39.1%;  Pred. No. 2.2e-46;
  Matches  326;  Conservative  148;  Mismatches  247;  Indels  112;  Gaps   16;

Qy         22 LRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQLKEA 81
              |: ||:||: || : :    |:|:|||::|||| |: : |||| |:  | |::  :|| 
Db         18 LQMEVERLTRELDQVSSASAQSAQYGLSLLEEKSALQQKCEELETLYDNTRHELDITQEA 77

Qy         82 FGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENER 141
               :  |: |     |  :|::|: |||::|     :: :|: |||||: |   ::| :| 
Db         78 LTKFQTSQKVTNKTGIEQEDALLNESAARETSLNLQIFDLENELKQLRHELERVRNERDR 137

Qy        142 LASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQN 201
              :    :      : |  | ||: ::|: |||| |:| :|||||||||||||||| || : 
Db        138 MLQENSDFGRDKSDSEADRLRLKSELKDLKFRETRMLSEYSELEEENISLQKQVSSLRSS 197

Qy        202 QVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKEL 261
              |||||| ||||:|| || | || |:::   ||:|:|:|:|||||||: ||| | :|:||| 
Db        198 QVEFEGAKHEIRRLTEEVELLNQQVDELANLKKIAEKQMEEALETLQGEREAKYALKKEL 257

Qy        262 SHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPK 321
                :::    :: |:|  |:       :| |  |:| |      |:  | :|  |  | 
Db        258 DGHLNRESMYHISNLAYSI--RSNMEDNASNNSDGEE-----ENLALKRLEADLSTELKS 310

Qy        322 KEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLS 381
              :|          || ||:::::|::||::||  || ||  | | |:: |  |: ::  | 
Db        311 PDGTK-------CDLFSEIHLNELKKLEKQLESMESEKTHLTANLREAQTSLDKSQNELQ 363

Qy        382 EQQEKVTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHV 441
                  ::  |  :: || :|           :|  |  | |
Db        364 NFMSRLALLAAHVDALVQL-----------KKQIDVKEQG------------------ 392

Qy        442 AVAEAG----ELREQLKALRS------THEAREAQHAEEKGRYEAEGQALTEKVSLL-EK 490
                | |    || :||:|| |      |  |:|    :       :|   |:  : |  :
Db        393 --KEGGQKKDELEQQLRALISQYANWFTLSAKEIDGLKTDIAELQKGLNYTDATTTLRNE 450

Qy        491 ASRQDRELLARLEKELKKVSDV---------AGETQGSLSVAQDELVTFSEELANLYHHV 541
                :    :|||  :| |   |||         ||:: ||   |:  ||  |::|| ||| |
Db        451 VTNLKNKLLATEQKSLDLQSDVQTLTHISQNAGQSLGS---ARSTLVALSDDLAQLYHLV 507

Qy        542 CMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGG 601
              |  | ||| ||:||:  :          |: ::: :    :    ::    | ||
Db        508 CTVNGETPTRVLLDHKTDDMS----FENDSLTAIQSQFKSDVFIAKPQIVEDLQGLADS- 562

Qy        602 TGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQEL 661
                                  : |   :  : ||||:|: ||: | :::: :| | |
Db        563 ----------------------VEIKKYVDTVSDQIKYLKTAVEHTIDMNKHKIRS-EG 598

Qy        662 GPAVDK----DKEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYEN 717
              | |::|    : | | |:|:|||||| |||| ||| |||:|||||||| ||||||| 
Db        599 GDALEKVNTEEMEELQEQIVKLKSLLSVKREQIGTLRNVLKSNKQTAEVALTNLKSKYEN 658

Qy        718 EKAMVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKT 777
              || :|::|| ||||||: |||||||||||||||| ||:||:||:|:: ||| |||:|||| 
Db        659 EKIIVSDTMSKLRNELRLLKEDAATFSSLRAMFAARCEEYVTQVDDLNRQLEAAEEEKKT 718

Qy        778 LNSLLRMAIQQKLALTQRLELLELDHEQ--------TRRGRA-KAAPKTKPAT 821
              || |||:|:|||||||||| :|:| |           :|| : |::  |:|::
Db        719 LNQLLRLAVQQKLALTQRLEEMEMDREMRHVRRPMPAQRGTSGKSSFSTRPSS 771


RESULT 2
MYS_AEQIR
ID   MYS_AEQIR      STANDARD;      PRT;  1938 AA.
AC   P24733;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, striated muscle.
OS   Aequipecten irradians (Bay scallop).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pectinoida;
OC   Pectinoidea; Pectinidae; Argopecten.
OX   NCBI_TaxID=31199;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adductor muscle;
RX   MEDLINE=92011595; PubMed=1917970;
RA   Nyitray L., Goodwin E.B., Szent-Gyoergyi A.G.;
RT   "Complete primary structure of a scallop striated muscle myosin heavy
RT   chain. Sequence comparison with other heavy chains reveals regions
RT   that might be critical for regulation.";
RL   J. Biol. Chem. 266:18469-18476(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adductor muscle;
RX   MEDLINE=91088319; PubMed=2263488;
RA   Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT   "Nucleotide sequence of full length cDNA for a scallop striated
RT   muscle myosin heavy chain.";
RL   Nucleic Acids Res. 18:7158-7158(1990).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RX   MEDLINE=94173332; PubMed=8127365;
RA   Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
RA   Szent-Gyorgyi A.G., Cohen C.;
RT   "Structure of the regulatory domain of scallop myosin at 2.8-A
RT   resolution.";
RL   Nature 368:306-312(1994).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
```

```
RX   MEDLINE=96419133; PubMed=8805510;
RA   Houdusse A., Cohen C.;
RT   "Structure of the regulatory domain of scallop myosin at 2-A
RT   resolution: implications for regulation.";
RL   Structure 4:21-32(1996).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55714; CAA39247.1; -.
DR   PIR; S13557; S13557.
DR   PIR; A40997; A40997.
DR   PDB; 1SCM; 30-APR-94.
DR   PDB; 1WDC; 11-JUL-96.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Alkylation; Calmodulin-binding; 3D-structure.
FT   DOMAIN       1    777       MYOSIN HEAD-LIKE.
FT   DOMAIN     778    805       IQ.
FT   DOMAIN     836   1938       RODLIKE TAIL (S2 AND LMM DOMAINS).
FT   DOMAIN     836   1938       COILED COIL (POTENTIAL).
FT   NP_BIND    176    183       ATP (BY SIMILARITY).
FT   MOD_RES    693    693       ALKYLATION (SH-1) (BY SIMILARITY).
FT   MOD_RES    703    703       ALKYLATION (SH-2) (BY SIMILARITY).
FT   HELIX      778    821
FT   TURN       822    823
FT   HELIX      825    833
FT   TURN       834    835
SQ   SEQUENCE   1938 AA;  222821 MW;  A5CCE4127D1A4896 CRC64;

  Query Match           8.8%;  Score 358.5;  DB 1;  Length 1938;
  Best Local Similarity  21.7%;  Pred. No. 2.5e-07;
  Matches  197;  Conservative  155;  Mismatches  352;  Indels  203;  Gaps   32;

Qy         13 LVMEAQPEWLRAEVK---RLSHELAETTREKIQAAEYGLAVLEEKHQLKLQFEELE---- 65
Db        837 LSIARQEEEMKEQLKQMDKMKEDLAKTERIKKELEEQNVTLLEQKNDLFLQLQTLEDSMG 896

Qy         66 ---------------------------------VDYEAIRSEME----QLKEAFGQAHT 87
Db        897 DQEERVEKLIMQKADFESQIKELEERLLDEEDAAADLEGIKKKMEADNANLKKDIGDLEN 956

Qy         88 NHKKVAADGESREE--SLIQESASKEQYYVRKVLELQTELKQLRNVLTNT----QSE--- 138
Db        957 TLQKAEQDKAHKDNQISTLQGEISQQDEHIGK---LNKEKKALEEANKKTSDSLQAEEDK 1013

Qy        139 ----NERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQ 194
Db       1014 CNHLNKLKAKLEQALDELEDNLEREK-KVRGDVEKAK----------RKVEQDLKSTQEN 1062

Qy        195 VSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRL-----KEISERQ--LEEALETL 247
Db       1063 VEDLERVKRE---LEENVRRKEAEISSLNSKLEDEQNLVSQLQRKIKELQARIEELEEEL 1119

Qy        248 KTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGG 307
Db       1120 EAERNARAKVEKQ------------RAELNRELEELGERLDEAGGATSAQIELNKKREAE 1167

Qy        308 LAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSEL--NISEIQKLKQQLMQMEREKAGLLAT 365
Db       1168 LLKIRRDLEEASLQHEAQISALRKKHQDAANEMADQVDQLQKVKSKL---EKDKKDLKRE 1224

Qy        366 LQDTQKQLEHTRGSLSEQQEKVTRLTENLSALR-RLQASKERQTALDNEKDRDSHEDGDY 424
Db       1225 MDDLESQMTHNMKNKGCSEKVMKQFESQMSDLNARLEDSQRSINELQSQKSRLQAENSDL 1284

Qy        425 YEVDINGPEILACKYHVAV--AEAGELREQLKALRSTHEAREAQHAEEKGRYEAEGQALT 482
Db       1285 TR------QLEDAEHRVSVLSKEKSQLSSQLE------DARRSLEEETRARSKLQNEVRN 1332

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDEL----VTFSEELANLY 538
Db       1333 MHADM--DAIRE------QLEEEQESKSDV----QRQLSKANNEIQQWRSKFESEGANRT 1380

Qy        539 HHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPE---- 594
Db       1381 EEL-----EDQKRKLL----------GKLSEAEQTTEAANAKCSALEKAKSRLQQELEDM 1425

Qy        595 ----------------AGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQI 638
Db       1426 SIEVDRANASVNQMEKKQRAFDKTTAEWQAKVNSLQSELENSQKESRGYSAELYRIKASI 1485

Qy        639 KHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLS---TKREQITTLRT 695
Db       1486 EEYQDSIG----------------ALRRENKNLADEIHDLTDQLSEGGRSTHELDKARR 1528

Qy        696 VLKANKQTAEVALANLKSKYENEKAMVTETMMKL---RNEL-KALKEDAATFSSLRAMFA 751
Db       1529 RLEMEKEELQAALEEAEGALEQEEAKVMRAQLEIATVRNEIDKRIQEKEEEFDNTRR--- 1585

Qy        752 TRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRA 811
Db       1586 ----NHQRALESMQASLEAEAKGKAD------AMRIKKKLEQDINELEVALDASNRGKA 1634

Qy        812 KAAPKTK 818
Db       1635 EMEKTVK 1641


RESULT 3
USO1_YEAST
ID   USO1_YEAST     STANDARD;      PRT;  1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Intracellular protein transport protein USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
```

```
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC       MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC       ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC       OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC       COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC   -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   PIR; A38455; A38455.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR002017; Spectrin.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT   DOMAIN        1    724       GLOBULAR HEAD.
FT   DOMAIN      725   1790       COILED COIL (POTENTIAL).
FT   DOMAIN      465    487       CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN      991   1790       DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN     1172   1786       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    847    847       G -> E (IN REF. 2).
FT   CONFLICT    924    924       E -> K (IN REF. 2).
FT   CONFLICT   1253   1253       V -> I (IN REF. 2).
FT   CONFLICT   1319   1319       I -> V (IN REF. 2).
FT   CONFLICT   1461   1461       N -> S (IN REF. 2).
FT   CONFLICT   1581   1581       G -> S (IN REF. 2).
FT   CONFLICT   1600   1600       I -> V (IN REF. 2).
FT   CONFLICT   1661   1661       R -> S (IN REF. 2).
FT   CONFLICT   1772   1772       D -> DEEDDEE (IN REF. 2).
SQ   SEQUENCE   1790 AA;  206424 MW;  6CE2B216E9FD4818 CRC64;

  Query Match             8.6%;  Score 352.5;  DB 1;  Length 1790;
  Best Local Similarity   20.5%;  Pred. No. 3.9e-07;
  Matches  193;  Conservative  176;  Mismatches  307;  Indels  266;  Gaps   34;

Qy         25 EVKRLSHELAETTREKIQAAEYGL-AVLEEKHQLKLQFEELEVDYEAIRSEME------- 76
              | |   |:      : |:  | ||  :| :| : :    ::  |  |:  ||:
Db        829 EYKSTIHK----QEDSIKTLEKGLETILSQKKKAEDGINKMGKDLFALSREMQAVEENCK 884

Qy         77 QLKEAFGQAHTNHKK--------------VAADGESREESLIQ-ESASKEQYYVRK--- 117
               |::   ::: ||:|               : |  |: ||  ||  : |||: :: |
Db        885 NLQKEKDKSNVNHQKETKSLKEDIAAKITEIKAINENLEEMKIQCNNLSKEKEHISKELV 944

Qy        118 ------------VLELQTELKQLRNVLTNTQSENERLASVAQELK-------------- 150
                        | :|  :|| | |   : |:||| |    :| |
Db        945 EYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAENESLIKAVEESKNESSIQLSNLQNKID 1004

Qy        151 ---EINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENIS--------LQKQVSVLR 199
                 :  :| :|:|| :  :|:: |    : : |  : :|| ||         : |:|:|:
Db       1005 SMSQEKENFQIERGSIEKNIEQLK----KTISDLEQTKEEIISKSDSSKDEYESQISLLK 1060

Qy        200 Q-------------NQVE--------------FEGLKHEIK-RLEEETEYLNSQLEDAI 230
```

```
                :              |::                  :: ||:|:: :||   : |    |:
Db       1061 EKLETATTANDENVNKISELTKTREELEAELAAYKNLKNELETKLETSEKALKEVKENEE 1120

Qy        231 RLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAA 290
              |||   :  :|| ||    ::| |||| |       :|    |  ::    |: :  |
Db       1121 HLKEEKIQLEKEATET----KQQLNSLRANL-------ESLEKEHEDLAAQLKKYEEQIA 1169

Qy        291 EPN---NDAEALVNGFEHGGLAKLPLDNKTSTPKK-EGLAPPSPSL---------VSDLL 337
                    |:  : :|           |  ||| :: | :   :  |          |:
Db       1170 NKERQYNEEISQLN-----------DEITSTQQENESIKKKNDELEGEVKAMKSTSEEQ 1217

Qy        338 SELNISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLSAL 397
              | |  |||  |  |: :::::     |:| :: | :|     : | |::     : :| |
Db       1218 SNLKKSEIDALNLQIKELKKKNETNEASLLESIKSVESETVKIKELQDECNFKEKEVSEL 1277

Qy        398 R-RLQASKERQTA-LDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKA 455
               :|:||::: :  |: :|: :  ::    |:|                |:  ||: ||:
Db       1278 EDKLKASEDKNSKYLELQKESEKIKE----ELD---------------AKTTELKIQLEK 1318

Qy        456 LRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDR----ELLARLEKELKKVSD 511
              : :  :|:|   :|                :| |:| | ::|    | | :|: |: :: :
Db       1319 ITNLSKAKEKSESE--------------LSRLKKTSSEERKNAEEQLEKLKNEI-QIKN 1362

Qy        512 VAGE------TQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAG 565
              | |       :|| :: |:    :||::  |   :    ||   :
Db       1363 QAFEKERKLLNEGSSTITQE----YSEKINTLEDELIRLQNENELK-------------- 1404

Qy        566 RTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPM 625
                    :   |    : |   |  |                 ||   :
Db       1405 -----AKEIDNTRSELEKVSLSNDELLEEKQNTIKSLQDEI----------LSYKDKITR 1449

Qy        626 NIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLST 685
              |   |::| ||  : |::       |  |  |:||    |::  :  | ||  | |: |
Db       1450 NDEKLLSIERDNKRDLES------LKEQLRAAQESKAKVEEGLKKLEEESSKEKAELEK 1502

Qy        686 KREQITTLRTVLKAN--------------------KQTAEVALANLK-------SKYEN 717
              :| :  | : :::|                    |::||  : ||:       |:
Db       1503 SKEMMKKLESTIESNETELKSSMETIRKSDEKLEQSKKSAEEDIKNLQHEKSDLISRINE 1562

Qy        718 EKAMVTETMMKLR------NELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAA 771
              :  : |   |||      :||: :|::          : |       ::|::::|:|
Db       1563 SEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINAEENTVLKSKLEDIERELKDK 1622

Qy        772 EDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKA 813
              : | |:          ::|  || ||: || : : |::   |:
Db       1623 QAEIKSNQ-------EEKELLTSRLKELEQELDSTQQKAQKS 1657
```

```
RESULT 4
MYHB_HUMAN
ID   MYHB_HUMAN     STANDARD;      PRT;  1972 AA.
AC   P35749; O00396; P78422; O94944;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, smooth muscle isoform (SMMHC).
GN   MYH11 OR KIAA0866.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99425270; PubMed=10493829;
RA   Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA   Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA   Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA   Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT   "Genome duplications and other features in 12 Mb of DNA sequence from
RT   human chromosome 16p and 16q.";
RL   Genomics 60:295-308(1999).
RN   [2]
RP   SEQUENCE OF 1-1266 FROM N.A.
```

```
RC   TISSUE=Brain;
RX   MEDLINE=99156230; PubMed=10048485;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:355-364(1998).
RN   [3]
RP   SEQUENCE OF 885-1972 FROM N.A.
RX   MEDLINE=93263189; PubMed=7684189;
RA   Matsuoka R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
RA   Yanagisawa M., Masaki T., Takao A.;
RT   "Human smooth muscle myosin heavy chain gene mapped to chromosomal
RT   region 16q12.";
RL   Am. J. Med. Genet. 46:61-67(1993).
RN   [4]
RP   SEQUENCE OF 1093-1972 FROM N.A.
RC   TISSUE=Hippocampus;
RA   Okajima K.;
RL   Submitted (NOV-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
CC       ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- DISEASE: A chromosomal rearrangement, known as pericentric
CC       inversion Inv(16)(p13q22), Produces a fusion protein that consists
CC       of the 165 N-terminal residues of CBF-beta (PEPB2) with the tail
CC       region of MYH11. This rearrangement is associated with acute
CC       myeloid leukemia of M4EO subtype.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF001548; AAC31665.1; -.
DR   EMBL; U91323; AAC35212.1; -.
DR   EMBL; AB020673; BAA74889.1; -.
DR   EMBL; D10667; -; NOT_ANNOTATED_CDS.
DR   EMBL; X69292; CAA49154.1; -.
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7569; MYH11.
DR   MIM; 160745; -.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 2.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW   Multigene family; Proto-oncogene; Chromosomal translocation.
FT   DOMAIN        1    785       MYOSIN HEAD-LIKE.
FT   DOMAIN      786    815       IQ.
FT   DOMAIN      844   1934       COILED COIL (POTENTIAL).
FT   DOMAIN     1935   1972       CARBOXYL-TERMINAL.
FT   NP_BIND     178    185       ATP (POTENTIAL).
FT   DOMAIN      661    683       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN      762    776       ACTIN-BINDING (BY SIMILARITY).
FT   MOD_RES     129    129       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     711    711       ALKYLATION (SH-2) (POTENTIAL).
FT   CONFLICT    887    889       EEK -> NSE (IN REF. 3).
FT   CONFLICT   1263   1266       ELQS -> TLSF (IN REF. 2).
FT   CONFLICT   1558   1558       T -> S (IN REF. 3).
FT   CONFLICT   1610   1611       KQ -> NE (IN REF. 3).
FT   CONFLICT   1786   1786       A -> S (IN REF. 4).
FT   CONFLICT   1958   1958       T -> L (IN REF. 3).
SQ   SEQUENCE   1972 AA;  227338 MW;  67665BB2AECE1277 CRC64;

  Query Match             8.6%;  Score 352;  DB 1;  Length 1972;
  Best Local Similarity   20.9%;  Pred. No. 4.5e-07;
  Matches  187;  Conservative  176;  Mismatches  342;  Indels  188;  Gaps   31;

Qy          9 EYARLVMEAQPEWLRAEVKRLSHEL------AETTREKIQAAEYGLAVLEEKHQLKLQFE 62
              :: ||  : :|     :| |   |:       : |:|: | ||  |  ||:||
Db        833 QWWRLFTKVKP---LLQVTRQEEEMQAKEDELQKTKERQQKAENELKELEQKH------S 883

Qy         63 ELEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQ 122
              :|  :   :: :::   | : :|     ::||  :  || | :  |  |:   |  :||
Db        884 QLTEEKNLLQEQLQAETELYAEAEEMRVRLAAKKQELEEILHEMEARLEEEEDRG-QQLQ 942

Qy        123 TELKQLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD-- 180
               | |::   : : : : |   :  |:|:      | :  :| |:|         |: |
Db        943 AERKKMAQQMLDLEEQLEEEEAARQKLQLEKVTAEAKIKKLEDEI---------LVMDDQ 993

Qy        181 YSELEEENISLQKQVSVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQL 240
              ::| :|   |::::| |  |  | |     : :|: : | : |:||  :: :| | ::|
Db        994 NNKLSKERKLLEERISDLTTNLAEEEEKAKNLTKLKNKHESMISELEVRLKKEEKSRQEL 1053

Qy        241 EEALETLKTE----REQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN-- 294
              |:    |: :     ||   |: ::: : :  :      | :|   :  |: |: ||
Db       1054 EKLKRKLEGDASDFHEQIADLQAQIAE-LKMQLAKKEEELQAAL--ARLDDEIAQKNNAL 1110

Qy        295 ----DAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSELNISEIQKLKQ 350
                  : |  ::  :    ::   ||    |:: |     :| ::|   |: : |  |
Db       1111 KKIRELEGHISDLQEDLDSERAARNKAEKQKRD-LGEELEALKTELEDTLDSTATQ---Q 1166

Qy        351 QLMQMEREKAGLLATLQDTQKQLEHTRGSLSEQQE-------KVTRLTENLSALRRLQAS 403
              :|     ::  :|    |     | ||   :: ||         |  ||| |   :| :|:
Db       1167 ELRAKREQEVTVLKKALD-----EETRSHEAQVQEMRQKHAQAVEELTEQLEQFKRAKAN 1221

Qy        404 KERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEAR 463
                 ||  |    |: |                      |||||   :|     :
Db       1222 ------LDKNKQTLEKENADL--------------------AGELRVLGQA------KQ 1248

Qy        464 EAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLARLEKELKKVSDVAGETQGSLSVA 523
               | :|  :| : ||: | |  | |  |:|  :  : : :|: |:: |: :  | :|
Db       1249 EVEH--KKKKLEAQVQELQSKCSDGERARAELNDKVHKLQNEVESVTGMLNEAEGKAIKL 1306

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
               :: : | :| :       :   ||  :: :   :  |    | |   :   |   :::
Db       1307 AKDVASLSSQLQDTQE---LLQEETRQKLNVS-TKLRQLEEERNSLQDQLDEEMEAKQN- 1361

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQA 643
               | : :                    |:|   ||| :::  :  : : :  :  |  |
Db       1362 --LERHI------------------STLNIQLSDSKKKLQDFASTVEALEEGKKRFQK 1399

Qy        644 AVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTL-RTVLKANKQ 702
               ::  |:   :: |: :    ::| |  | :|:  |   |  :|: :: | :   | ::
```

```
Db       1400 EIENLTQQYEEKAAAYD---KLEKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQRKFDQL 1456

Qy        703 TAEVALANLKSKYENEK----------------AMVTETMMKLRNE-------LKALKE 738
              ||    |: ||| :|:                 |   |  :: : |       ||| |
Db       1457 LAEE--KNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELERTNKMLKAEME 1514

Qy        739 DAATF-----------SSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
              |  :             | ||:  |: :|  |||:|:: :| | || |      ||: :
Db       1515 DLVSSKDDVGKNVHELEKSKRAL-ETQMEEMKTQLEELEDELQATEDAK------LRLEV 1567

Qy        787 --------------------QQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
                                  :::  | ::|   | : |  |: || ||   |
Db       1568 NMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKK 1620


RESULT 5
MYS2_DICDI
ID   MYS2_DICDI     STANDARD;      PRT;  2116 AA.
AC   P08799;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin II heavy chain, non muscle.
GN   MHCA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87092266; PubMed=3540939;
RA   Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT   "Conserved protein domains in a myosin heavy chain gene from
RT   Dictyostelium discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN   [2]
RP   PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC   STRAIN=AX2;
RX   MEDLINE=90353583; PubMed=2387408;
RA   Lueck-Vielmeter D., Schleicher M., Grabatin B., Wippler J.,
RA   Gerisch G.;
RT   "Replacement of threonine residues by serine and alanine in a
RT   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL   FEBS Lett. 269:239-243(1990).
RN   [3]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=88112226; PubMed=2828113;
RA   Wagle G., Noegel A., Scheel J., Gerisch G.;
RT   "Phosphorylation of threonine residues on cloned fragments of the
RT   Dictyostelium myosin heavy chain.";
RL   FEBS Lett. 227:71-75(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345066; PubMed=7619795;
RA   Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA   Rayment I.;
RT   "X-ray structures of the myosin motor domain of Dictyostelium
RT   discoideum complexed with MgADP.BeFx and MgADP.AlF4-.";
RL   Biochemistry 34:8960-8972(1995).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX   MEDLINE=95345067; PubMed=7619796;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT   truncated head of Dictyostelium discoideum myosin to 2.7-A
RT   resolution.";
RL   Biochemistry 34:8973-8981(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=96206189; PubMed=8611530;
RA   Smith C.A., Rayment I.;
RT   "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL   Biochemistry 35:5404-5417(1996).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX   MEDLINE=97452580; PubMed=9305951;
RA   Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT   "X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes
RT   of the Dictyostelium discoideum myosin motor domain.";
RL   Biochemistry 36:11619-11628(1997).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX   MEDLINE=98070605; PubMed=9405148;
RA   Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT   "X-ray crystal structure and solution fluorescence characterization
RT   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT   Dictyostelium discoideum myosin motor domain.";
RL   J. Mol. Biol. 274:394-407(1997).
CC   -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC       ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC   -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC       INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC       LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC       (MLC-2).
CC   -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC       CORTEX.
CC   -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC       SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC       THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC       ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC       POSITION (688).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14628; AAA33227.1; -.
DR   PIR; A26655; A26655.
DR   PIR; S00250; S00250.
DR   PDB; 1MMA; 03-DEC-97.
DR   PDB; 1MMD; 17-AUG-96.
DR   PDB; 1MMG; 03-DEC-97.
DR   PDB; 1MMN; 03-DEC-97.
DR   PDB; 1MND; 17-AUG-96.
DR   PDB; 1MNE; 17-AUG-96.
DR   PDB; 1VOM; 23-DEC-96.
DR   PDB; 1LVK; 28-JAN-98.
DR   DictyDb; DD01008; mhcA.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW   Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT   DOMAIN        1    761       MYOSIN HEAD-LIKE.
FT   DOMAIN      762    791       IQ.
```

```
FT   DOMAIN      817   2116       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP.
FT   DOMAIN      638    660       ACTIN-BINDING.
FT   DOMAIN      738    752       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (DI-) (POTENTIAL).
FT   MOD_RES     678    678       ALKYLATION (SH-1).
FT   MOD_RES    1823   1823       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    1833   1833       PHOSPHORYLATION (BY MHCK).
FT   MOD_RES    2029   2029       PHOSPHORYLATION (BY MHCK).
SQ   SEQUENCE   2116 AA;  243871 MW;  2FC3770BB1EE56A1 CRC64;

  Query Match             8.5%;  Score 346.5;  DB 1;  Length 2116;
  Best Local Similarity   21.0%;  Pred. No. 8.1e-07;
  Matches  224;  Conservative  161;  Mismatches  338;  Indels  345;  Gaps   43;

Qy          6 EEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEK--HQLKLQFEE 63
              :||:  |  :| : : | | :: :  :|  |  ||   : | | | :|     |: |  |
Db       1104 DEEKKNRDALEKKKKALDAMLEEMKDQLESTGGEK--KSLYDLKVKQESDMEALRNQISE 1161

Qy         64 --------------LEVDYEAIRSEMEQLKEAFGQAHTNHKKVAADGESREESLIQESAS 109
                           || :   :: |:|  : |      |||  | | :   | :|:|:
Db       1162 LQSTIAKLEKIKSTLEGEVARLQGELEAEQLAKSNVEKQKKKVELDLEDKSAQLAEETAA 1221

Qy        110 KE------QYYVRKVLELQTELKQLRNVLTNTQSENERLAS------------------- 144
              |:      :   ::: |:||:| :  |   |: | |: | :
Db       1222 KQALDKLKKKLEQELSEVQTQLSEANNKNVNSDSTNKHLETSFNNLKLELEAEQKAKQAL 1281

Qy        145 ------VAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELE---EENISLQKQV 195
                   :  ||| :|: :| : : :   : |    |:  | :: |||:   || :: :| |
Db       1282 EKKRLGLESELKHVNEQLEEEKKQKESNEK----RKVDLEKEVSELKDQIEEEVASKKAV 1337

Qy        196 SVLR-QNQVEFEGLK--------------HEIKRLEEETEYLNS-------QLEDAIRLK 233
              :  : : : | : :|               ::| |: : | | :        ||: | | |
Db       1338 TEAKNKKESELDEIKRQYADVVSSRDKSVEQLKTLQAKNEELRNTAEEAEGQLDRAERSK 1397

Qy        234 EISERQLEEALETLKTEREQK--------------NSLRKELSHYMSINDSFYT--SHLH 277
              : :|  ||||:: |: | :|                | : ||    :::   |     |:
Db       1398 KKAEFDLEEAVKNLEEETAKKVKAEKAMKKAETDYRSTKSELDDAKNVSSEQYVQIKRLN 1457

Qy        278 VSLDGLK-FSDDAAEPNNDA-----------EALVNGFEHGGLAKLPLDNKTSTPKKEGL 325
               |  |:   ::| |  | |           |:| :  :    ||   : | |   :  :
Db       1458 EELSELRSVLEEADERCNSAIKAKKTAESALESLKDEIDAANNAKAKAERK-SKELEVRV 1516

Qy        326 APPSPSLVSDLLSELNISEIQKLKQQL----MQMEREKAGLLATLQD---TQKQLEHTRG 378
              |    ||  |    :|:  |:|   ::     :::||    : : :|   |:||
Db       1517 AELEESL-EDKSGTVNVEFIRKKDAEIDDLRARLDRETESRIKSDEDKKNTRKQFADLEA 1575

Qy        379 SLSEQQEKVT---RLTENLSA------------LRRLQASKERQTALDNEKDRDSHEDG 422
              : | | :|    || : | :               |::  | ::       :| : |:|
Db       1576 KVEEAQREVVTIDRLKKKLESDIIDLSTQLDTETKSRIKIEKSKKKLEQTLAERRAAEEG 1635

Qy        423 D------------YYEVD------------INGPEILACKYHVAVAEAGELREQL----- 453
                           : |||            :|  |    |    |||  |::|||
Db       1636 SSKAADEEIRKQVWQEVDELRAQLDSERAALNASE---KKIKSLVAEVDEVKEQLEDEIL 1692

Qy        454 ------KALRS----THEAREAQHAEE--------------------KGRYEAE------ 477
                    || |:      | |:   ||                      | :|:||
Db       1693 AKDKLVKAKRALEVELEEVRDQLEEEEDSRSELEDSKRRLTTEVEDIKKKYDAEVEQNTK 1752

Qy        478 ----GQALTEKVSLLEK-----------------ASRQDRELLARLEKELKKVS----- 510
                  : ||: |  |:|                   :: : ||:|: |:|  |
Db       1753 LDEAKKKLTDDVDTLKKQLEDEKKKLNESERAKKRLESENEDFLAKLDAEVKNRSRAEKD 1812

Qy        511 -----------------DVAGETQGSLSVAQ-----DEL-----------------VTF 530
                              : | :||  :  |:     |||                  |
Db       1813 RKKYEKDLKDTKYKLNDEAATKTQTEIGAAKLEDQIDELRSKLEQEQAKATQADKSKKTL 1872

Qy        531 SEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGL 590
              |: ||   :     |   :: :   :| :   |      |  ||   :|     | |
Db       1873 EGEIDNLRAQI-----EDEGKIKMRLEKEKRALEGELEELRETVEEAEDSKSEAEQSKRL 1927

Qy        591 LAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQIKHLQ-AAVDRTT 649
              :  |                      | | ||      :  |  |   :||   |:

Db       1928 VELE--------------------LEDARRNLQKEIDAKEIAEDAKSNLQREIVEAKG 1965

Qy        650 ELSRQRIASQELGPAVDKDKEALMEEILKLKSLL----STKREQITTLRTVLKANKQTAE 705
                  |  : ||         |: :: |  ||  | : :       | :||       | ||:  |
Db       1966 RLEEESIARTN----SDRSRKRLEAEIDALTAQVDAEQKAKNQQI-------KENKK-IE 2013

Qy        706 VALANLKSKY-ENEKAMVTETMM--KLRNELKALKEDAATFSSLRAMFATRCDEYITQLD 762
                 |   : |: |:||    | ::  ||  : |  |::||                  |
Db       2014 TELKEYRKKFGESEKTKTKEFLVVEKLETDYKRAKKEAA-------------------D 2053

Qy        763 EMQRQLAAAEDEKKTLN--SLLRMAIQQKLALTQRLELLELDHEQTRR 808
                | |::|    | :| |:  |||: || :          |: ||::|:|
Db       2054 EQQQRLTVENDLRKHLSEISLLKDAIDK----------LQRDHDKTKR 2091


RESULT 6
MYHA_HUMAN
ID   MYHA_HUMAN     STANDARD;      PRT;  1976 AA.
AC   P35580;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE   type B) (Nonmuscle myosin heavy chain-B) (NMMHC-B).
GN   MYH10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96025307; PubMed=7499478;
RA   Phillips C.L., Yamakawa K., Adelstein R.S.;
RT   "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and
RT   analysis of human tissues with isoform-specific antibodies.";
RL   J. Muscle Res. Cell Motil. 16:379-389(1995).
RN   [2]
RP   SEQUENCE OF 63-722 FROM N.A.
RX   MEDLINE=91316803; PubMed=1860190;
RA   Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA   Gdula D., Adelstein R.S., Weir L.;
RT   "Human nonmuscle myosin heavy chains are encoded by two genes located
RT   on different chromosomes.";
RL   Circ. Res. 69:530-539(1991).
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING.
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69181; AAA99177.1; -.
DR   PIR; B61231; B61231.
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7568; MYH10.
DR   MIM; 160776; -.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
```

```
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Alkylation; Multigene family.
FT   DOMAIN       1    785       MYOSIN HEAD-LIKE.
FT   DOMAIN     786    815       IQ.
FT   DOMAIN     845   1976       COILED COIL (POTENTIAL).
FT   NP_BIND    178    185       ATP (POTENTIAL).
FT   MOD_RES    701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES    711    711       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1976 AA;  228938 MW;  B2BB87FF35EA124F CRC64;

  Query Match           8.4%;  Score 346;  DB 1;  Length 1976;
  Best Local Similarity  21.8%;  Pred. No. 7.9e-07;
  Matches  207;  Conservative  170;  Mismatches  306;  Indels  268;  Gaps   39;

Qy          5 SEEEEYAR---LVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQF 61
              :|||| |:    :   |   :    :|| |  |.||:::: |:  |     |   |: |
Db       1016 AEEEEKAKNLAKIRNKQEVMISDLEERLKKE--EKTRQELEKAKRKLD--GETTDLQDQI 1071

Qy         62 EELEVDYEAIRSEMEQLKE------AFGQAHTNHK----KVAADGESR----EESLIQES 107
               ||:   : :: :: : :|      | |   | ||    ||  : :::    :|    |
Db       1072 AELQAQIDELKLQLAKKEEELQGALARGDDETLHKNNALKVVRELQAQIAELQEDFESEK 1131

Qy        108 ASKEQYYVRKVLELQTELK----QLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRL 163
              ||: :   :|  :|  ||:    :| : |  | :: |      ||: |: : :| :
Db       1132 ASRNKAEKQK-RDLSEELEALKTELEDTLDTTAAQQELRTKREQEVAELKKALEEETKNH 1190

Qy        164 RDDIKEYKFREARLLQDYSE-----------------LEEENISLQKQVSVLRQNQVEF 205
                |:: : | |  |:: ||                 || :|  |  :| ||:| : |
Db       1191 EAQIQDMRQRHATALEELSEQLEQAKRFKANLEKNKQGLETDNKELACEVKVLQQVKAES 1250

Qy        206 EGLKHEIKRLEEETEYLNSQLEDAIRLK-EISER--QLEEALE---TLKTEREQKNSLRK 259
              |   |: |:|: : : |:::: :  ||: |::|:  :|:  |:   ||  | |:|
Db       1251 E---HKRKKLDAQVQELHAKVSEGDRLRVELAEKASKLQNELDNVSTLLEEAEKK----- 1302

Qy        260 ELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNN---DAEALVNGFEHGGLAKLPLDNK 316
                                |:||: |||   :   | : |:                :
Db       1303 ----------------------GIKFAKDAASLESQLQDTQELL--------------Q 1325

Qy        317 TSTPKKEGLAPPSPSLVSDLLSELN-ISEIQKLKQQLMQ-MEREKAGLLATLQDTQKQLE 374
                | :|  |:    | :  |  | | : | |: :::  : :|::   | : | ||:|:::
Db       1326 EETRQKLNLS----SRIRQLEEEKNSLQEQQEEEEEARKNLEKQVLALQSQLADTKKKVD 1381

Qy        375 HTRGSLSEQQEKVTRLTENLSAL-RRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPE 433
              |::   :|   :| ::  || :||:         |:  |:|   |  |   ||::
Db       1382 DDLGTIESLEEAKKKLLKDAEALSQRLEEKALAYDKLEKTKNRLQQELDD-LTVDLDHQR 1440

Qy        434 ILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAE-----------GQALT 482
               :|       :  :|  : |::        |::|||: | |||            :||
Db       1441 QVASNLEKKQKKFDQLLAEEKSI-------SARYAEERDRAEAEAREKETKALSLARALE 1493

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVA------GETQGSLSVAQDELVTFSEEL-- 534
              | :   |:  ||:::| | :|  :    ||        ::: :|   ':|: |  |||
Db       1494 EALEAKEEFERQNKQLRADMEDLMSSKDDVGKNVHELEKSKRALEQQVEEMRTQLEELED 1553

Qy        535 -------ANLYHHVCMC---------------NNETPNRVMLDYYREGQGGAGRTSPGGR 572
                    | |   | |                  ||   |:::   ||
Db       1554 ELQATEDAKLRLEVNMQAMKAQFERDLQTRDEQNEEKKRLLIKQVRE------------- 1600

Qy        573 TSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE--------- 623
                                                 | : | | |::
Db       1601 ----------------------------------------LEAELEDERKQRALAVASKK 1620

Qy        624 --PMNIYNLIAII------RD----QIKHLQAAV-DRTTELSRQRIASQEL---GPAVDK 667
                  :::  :| | |      ||    |:: ||| : |   ||   | :  |:       :|
Db       1621 KMEIDLKDLEAQIEAANKARDEVIKQLRKLQAQMKDYQRELEEARASRDEIFAQSKESEK 1680

Qy        668 DKEALMEEILKL-KSLLSTKR------------EQITTL---RTVLKANKQTAEVALAN 710
                ::|  |||:| : | |::|              ::||     :: |   |:  |  :|
Db       1681 KLKSLEAEILQLQEELASSERARRHAEQERDELADEITNSASGKSALLDEKRRLEARIAQ 1740

Qy        711 LKSKYENEKA---MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQ 767
                |: : | |::   :: :   |   ::  |  : |   |     | : |    ||:   ::
Db       1741 LEEELEEEQSNMELLNDRFRKTTLQVDTLNAELAAERSA----AQKSDNARQQLERQNKE 1796

Qy        768 LAAAEDEKKTLNSLLRMAIQQKL-----ALTQRLELLELDHEQTRRGRAKA 813
                | |   |       |  |:: |      ||  ::  ||   ||  : || |
Db       1797 LKAKLQE-------LEGAVKSKFKATISALEAKIGQLEEQLEQEAKERAAA 1840


RESULT 7
MYHA_RAT
ID   MYHA_RAT       STANDARD;      PRT;  1976 AA.
AC   Q9JLT0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE   type B) (Nonmuscle myosin heavy chain-B) (NMMHC-B).
GN   MYH10.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=20483650; PubMed=11027611;
RA   Yam J.W.P., Chan K.W., Li N., Hsiao W.L.W.;
RT   "Molecular cloning and functional analysis of the promoter region of
RT   rat nonmuscle myosin heavy chain-B gene.";
RL   Biochem. Biophys. Res. Commun. 276:1203-1209(2000).
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING.
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF139055; AAF61445.1; -.
DR   HSSP; P10587; 1BR2.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
```

```
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Alkylation; Multigene family.
FT   DOMAIN        1    785       MYOSIN HEAD-LIKE.
FT   DOMAIN      786    815       IQ.
FT   DOMAIN      845   1976       COILED COIL (POTENTIAL).
FT   NP_BIND     178    185       ATP (POTENTIAL).
FT   MOD_RES     701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     711    711       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1976 AA;  228963 MW;  E32708BF9BF2B470 CRC64;

  Query Match             8.4%;  Score 345;  DB 1;  Length 1976;
  Best Local Similarity   21.6%;  Pred. No. 8.6e-07;
  Matches  205;  Conservative  174;  Mismatches  304;  Indels  268;  Gaps   38;

Qy          5 SEEEEYAR---LVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQF 61
              :|||| |:    :   |   :    :||  |  | ||:::: |:  |    |   |: |
Db       1016 AEEEEKAKNLAKIRNKQEVMISDLEERLKKE--EKTRQELEKAKRKLD--GETTDLQDQI 1071

Qy         62 EELEVDYEAIRSEMEQLKE------AFGQAHTNHK----KVAADGESR----EESLIQES 107
               ||:   : :: :: : :|      | |   | ||    ||| : :::    :|    |
Db       1072 AELQAQVDELKVQLTKKEEELQGALARGDDETLHKNNALKVARELQAQIAELQEDFESEK 1131

Qy        108 ASKEQYYVRKVLELQTELK----QLRNVLTNTQSENERLASVAQELKEINQNVEIQRGRL 163
              ||: :   :|  :|  ||:    :| : |  | :: |      ||: |: : :| :
Db       1132 ASRNKAEKQK-RDLSEELEALKTELEDTLDTTAAQQELRTKREQEVAELKKALEDETKNH 1190

Qy        164 RDDIKEYKFREARLLQDYSE-----------------LEEENISLQKQVSVLRQNQVEF 205
                |:: : | |  |:: ||                 || :|  |  :| ||:| : |
Db       1191 EAQIQDMRQRHATALEELSEQLEQAKRFKANLEKNKQGLETDNKELACEVKVLQQVKAES 1250

Qy        206 EGLKHEIKRLEEETEYLNSQLEDAIRLK-EISER--QLEEALE---TLKTEREQKNSLRK 259
              |   |: |:|: : : |:::: :  ||: |::|:  :|:  |:   ||  | |:|
Db       1251 E---HKRKKLDAQVQELHAKVSEGDRLRVELAEKANKLQNELDNVSTLLEEAEKK----- 1302

Qy        260 ELSHYMSINDSFYTSHLHVSLDGLKFSDDAA---EPNNDAEALVNGFEHGGLAKLPLDNK 316
                                |:||: |||       | : |:                :
Db       1303 ----------------------GMKFAKDAAGLESQLQDTQELL--------------Q 1325

Qy        317 TSTPKKEGLAPPSPSLVSDLLSELN-ISEIQKLKQQLMQ-MEREKAGLLATLQDTQKQLE 374
              | :|  |:    | :  |  | | : | |: :::  : :|::  |  : | ||:|:::
Db       1326 EETRQKLNLS----SRIRQLEEEKNSLQEQQEEEEEARKNLEKQVLALQSQLADTKKKVD 1381

Qy        375 HTRGSLSEQQEKVTRLTENLSAL-RRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPE 433
              |::   :|   :| ::: || :||:       |:  |:|   |  |   ||::
Db       1382 DDLGTIEGLEEAKKKLLKDVEALSQRLEEKVLAYDKLEKTKNRLQQELDD-LTVDLDHQR 1440

Qy        434 ILACKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAE----------GQALT 482
               :       |:   : :::   | :  :   |::|||: | |||           :||
Db       1441 QI-------VSNLEKKQKKFDQLLAEEKGISARYAEERDRAEAEAREKETKALSLARALE 1493

Qy        483 EKVSLLEKASRQDRELLARLEKELKKVSDVA------GETQGSLSVAQDELVTFSEEL-- 534
              | :   |:  ||:::| | :|  :    ||        ::: :|    :|: |  |||
Db       1494 EALEAKEEFERQNKQLRADMEDLMSSKDDVGKNVHELEKSKRALEQQVEEMRTQLEELED 1553

Qy        535 -------ANLYHHVCMC---------------NNETPNRVMLDYYREGQGGAGRTSPGGR 572
                    | |   | |                  ||   |::|   ||
Db       1554 ELQATEDAKLRLEVNMQAMKAQFERDLQTRDEQNEEKKRLLLKQVRE------------- 1600

Qy        573 TSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE--------- 623
                                                 | : | | |::
Db       1601 ----------------------------------------LEAELEDERKQRALAVASKK 1620

Qy        624 --PMNIYNLIAII------RD----QIKHLQAAV-DRTTELSRQRIASQEL---GPAVDK 667
                 ::: :| | |      ||    |:: ||| : |   ||   | :  |:       :|
Db       1621 KMEIDLKDLEAQIEAANKARDEVIKQLRKLQAQMKDYQRELEEARASRDEIFAQSKESEK 1680

Qy        668 DKEALMEEILKLKSLLST----------KREQI-------TTLRTVLKANKQTAEVALAN 710
               ::|  |||:|:  |::          :|:::        : :: |   |:  |  :|
Db       1681 KLKSLEAEILQLQEELASSERARRHAEQERDELADEIANSASGKSALLDEKRRLEARIAQ 1740

Qy        711 LKSKYENEKA---MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQ 767
              |: : | |::   :: :   |   ::  |  : |   |     | : |    ||:   ::
```

```
Db       1741 LEEELEEEQSNMELLNDRFRKTTLQVDTLNTELAAERSA----AQKSDNARQQLERQNKE 1796

Qy        768 LAAAEDEKKTLNSLLRMAIQQKL-----ALTQRLELLELDHEQTRRGRAKA 813
                | |   |       |  |:: |      ||  ::  ||   ||  : || |
Db       1797 LKAKLQE-------LEGAVKSKFKATISALEAKIGQLEEQLEQEAKERAAA 1840


RESULT 8
MYHB_RABIT
ID   MYHB_RABIT     STANDARD;      PRT;  1972 AA.
AC   P35748;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, smooth muscle isoform (SMMHC).
GN   MYH11.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92073350; PubMed=1961735;
RA   Babij P., Kelly C., Periasamy M.;
RT   "Characterization of a mammalian smooth muscle myosin heavy-chain
RT   gene: complete nucleotide and protein coding sequence and analysis of
RT   the 5' end of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77812; AAA31395.1; -.
DR   PIR; A41604; A41604.
DR   HSSP; P10587; 1BR2.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW   Multigene family.
```

```
FT   DOMAIN        1    784       MYOSIN HEAD-LIKE.
FT   DOMAIN      785    807       IQ.
FT   DOMAIN      844   1934       COILED COIL (POTENTIAL).
FT   DOMAIN     1935   1972       CARBOXYL-TERMINAL.
FT   NP_BIND     178    185       ATP (POTENTIAL).
FT   DOMAIN      661    683       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN      763    777       ACTIN-BINDING (BY SIMILARITY).
FT   MOD_RES     129    129       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     711    711       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1972 AA;  227318 MW;  2061A224288D6A4C CRC64;

  Query Match             8.4%;  Score 344.5;  DB 1;  Length 1972;
  Best Local Similarity   21.8%;  Pred. No. 9e-07;
  Matches  200;  Conservative  157;  Mismatches  312;  Indels  249;  Gaps   37;

Qy         20 EWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEEKHQLKLQFEELEVDYEAIRSEMEQ 77
              :| |   |     | : ||  |::|| |  |  ::|: |                ||::: 
Db        833 QWWRLFTK--VKPLLQVTRQEEEMQAKEDELQKIKERQQ------------KAESELQE 877

Qy         78 LKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVLELQTELKQLRNVLTNTQS 137
              |:    | ||          | |::|:||    |        ||  | :::|      
Db        878 LQ----QKHTQ--------LSEEKNLLQEQLQAE-------TELYAEAEEMR-------- 910

Qy        138 ENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSV 197
                |||:  |||:||   :| :    |  :: :    :: |   :|||:   |::: :  
Db        911 --VRLAAKKQELEEILHEMEARLEEEEDRGQQLQAERKKMAQQMLDLEEQ---LEEEEAA 965

Qy        198 LRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSL 257
              :: |:|    : :||:||::   :: |     : ::: | :: :    |  | |:  :| 
Db        966 RQKLQLEKVTAEAKIKKLEDDILVMDDQNNKLSKERKLLEERISDLTTNLAEEEEKAKNL 1025

Qy        258 RKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKT 317
              |                        ||          |  |::::  |     :|  : | 
Db       1026 TK----------------------LK---------NKHESMISELE----VRLKKEEK- 1048

Qy        318 STPKKEGLAPPSPSLVSDL---LSELNISEIQKLKQQLMQMEREKAGLLATLQD------ 368
              |  : | |        |||   :::|  ::| :|| || : | |    || |:|      
Db       1049 SRQELEKLKRKMDGEASDLHEQIADLQ-AQIAELKMQLAKKEEELQAALARLEDETSQKN 1107

Qy        369 -TQKQLEHTRGSLSEQQEKVTR--------------LTENLSALR--------------R 399
               |::    | :|: || :                 | | | ||:               
Db       1108 NALKKIRELEGHISDLQEDLDSERAARNKAEKQKRDLGEELEALKTELEDTLDTTATQQE 1167

Qy        400 LQASKERQT-----ALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAEAGELREQLK 454
              |:| :|::       ||| |   |||     |:          : |  | |  || |||: 
Db       1168 LRAKREQEVTVLKKALDEE--TRSHE-AQVQEMR---------QKHTQVVE--ELTEQLE 1213

Qy        455 A-------LRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDREL--------- 498
                    |  | :  | ::|:  |     |||  |     :|   | :||            
Db       1214 QFKRAKANLDKTKQTLEKENADLAGELRVLGQAKQEVEHKKKKLEVQLQELQSKCSDGER 1273

Qy        499 -LARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVC----MCNNETPNRVM 553
               | |  :: |: :      | || |: : :  ::|:|:|   :     :   ||  ::   
Db       1274 ARAELNDKVHKLQNEVESVTGMLSEAEGKAIKLAKEVASLGSQLQDTQELLQEETRQKLN 1333

Qy        554 LDYYREGQGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSL 613
              :   :  |    | |   :   |  :::   | : :                     |:| 
Db       1334 VS-TKLRQLEDERNSLQEQLDEEMEAKQN---LERHI-------------------STL 1369

Qy        614 PSPLSDPRREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALM 673
                 ||| :::  :  : :  :  |  |  ::  |:   :: |: :    ::| |  |  
Db       1370 NIQLSDSKKKLQDFASTVESLEEGKKRFQKEIESLTQQYEEKAAAYD---KLEKTKNRLQ 1426

Qy        674 EEILKLKSLLSTKREQITTLRTVLKANKQTAEVAL--ANLKSKYENEK------------ 719
              :|:  |   |  :|: :: |    |  |:  ::     |: ||| :|:            
Db       1427 QELDDLVVDLDNQRQLVSNLE---KKQKKFDQLLAEEKNISSKYADERDRAEAEAREKET 1483

Qy        720 -----AMVTETMMKLRNE-------LKALKEDAATF------------SSLRAMFATRCD 755
                  |   |  :: : |       |||  ||. :              | ||:  |: :  
Db       1484 KALSLARALEEALEAKEELERTNKMLKAEMEDLVSSKDDVGKNVHELEKSKRAL-ETQME 1542

Qy        756 EYITQLDEMQRQLAAAEDEKKTLN---SLLRMAIQQKL-----------ALTQRLELLE 800
```

```
                |  |||:|:: :| | || |  |      |::  :: |             | ::|   |
Db       1543 EMKTQLEELEDELQATEDAKLRLEVNMQALKVQFERDLQARDEQNEEKRRQLQRQLHEYE 1602

Qy        801 LDHEQTRRGRAKAAPKTK 818
                : |  |: || ||   |
Db       1603 TELEDERKQRALAAAAKK 1620


RESULT 9
MYHA_BOVIN
ID   MYHA_BOVIN     STANDARD;      PRT;  1976 AA.
AC   Q27991;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE   type B) (Nonmuscle myosin heavy chain-B) (NMMHC-B).
GN   MYH10.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohara M., Ishiguro N., Shinagawa M.;
RT   "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 204-302 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=95301542; PubMed=7782316;
RA   Itoh K., Adelstein R.S.;
RT   "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle
RT   myosin heavy chain II-B.";
RL   J. Biol. Chem. 270:14533-14540(1995).
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING (BY SIMILARITY).
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB022023; BAA36494.1; -.
DR   EMBL; U15716; AAA87715.1; -.
DR   HSSP; P10587; 1BR2.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
```

```
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Alkylation; Multigene family.
FT   DOMAIN       1    785       MYOSIN HEAD-LIKE.
FT   DOMAIN     786    815       IQ.
FT   DOMAIN     845   1976       COILED COIL (POTENTIAL).
FT   NP_BIND    178    185       ATP (POTENTIAL).
FT   MOD_RES    701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES    711    711       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1976 AA;  229097 MW;  6144354451C0F790 CRC64;

  Query Match             8.3%;  Score 341;  DB 1;  Length 1976;
  Best Local Similarity   21.0%;  Pred. No. 1.2e-06;
  Matches  199;  Conservative  178;  Mismatches  309;  Indels  262;  Gaps   36;

Qy          5 SEEEEYAR---LVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGL-----AVLEEKHQ 56
              :|||| |:    :   |   :    :||  |  | ||:::: |:  |       : ::  :
Db       1016 AEEEEKAKNLAKIRNKQEVMISDLEERLKKE--EKTRQELEKAKRKLDGETTDLQDQIAE 1073

Qy         57 LKLQFEELEVDYEAIRSEMEQLKEAF--GQAHTNHK----KVAADGESR----EESLIQE 106
              |: | :||::    :  : |:|: |   |   | ||    ||  : :::    :|    |
Db       1074 LQAQIDELKIQ---VAKKEEELQGALARGDDETLHKNNALKVVRELQAQIAELQEDFESE 1130

Qy        107 SASKEQYYVRKVLELQTELK----QLRNVLTNTQSENERLASVAQELKEINQNVEIQRGR 162
               ||: :   :|  :|  ||:    :| : |  | :: |      ||: |: : :| :
Db       1131 KASRNKAEKQK-RDLSEELEALKTELEDTLDTTAAQQELRTKREQEVAELKKALEEETKS 1189

Qy        163 LRDDIKEYKFREARLLQDYSE-----------------LEEENISLQKQVSVLRQNQVE 204
                |:: : | |  |:: ||                  || :|  |  :| ||:| : |
Db       1190 HEAQIQDMRQRHATALEELSEQLEQAKRFKANLEKNKQGLETDNKELACEVKVLQQVKAE 1249

Qy        205 FEGLKHEIKRLEEETEYLNSQLEDAIRLK-EISER--QLEEALE---TLKTEREQKNSLR 258
               |  |: |:|: : : |:::: :  ||: |::|:  :|:  |:   ||  | |:|
Db       1250 SE---HKRKKLDAQVQELHAKVSEGDRLRVELAEKANKLQNELDNVSTLLEEAEKK---- 1302

Qy        259 KELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTS 318
                                 |:||: |||   :  :     :     || | ::
Db       1303 ----------------------GIKFAKDAAGLESQLQDTQELLQEETRQKLNLSSRIR 1339

Qy        319 TPKKEGLAPPSPSLVSDLLSELNISEIQKLKQQLMQ-MEREKAGLLATLQDTQKQLEHTR 377
               ::|                 :: | |: :::  : :|::   | | | ||:|:::
Db       1340 QLEEE---------------RSSLQEQQEEEEEARRSLEKQLQALQAQLTDTKKKVDDDL 1384

Qy        378 GSLSEQQEKVTRLTENLSAL-RRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILA 436
               |::   :|   :| :::  | :||:        |:  | |   |  |   ||::    :
Db       1385 GTIENLEEAKKKLLKDVEVLSQRLEEKALAYDKLEKTKTRLQQELDDLL-VDLDHQRQI- 1442

Qy        437 CKYHVAVAEAGELREQLKALRSTHEAREAQHAEEKGRYEAE-----------GQALTEKV 485
                 |:   : :::   | :  :   |::|||: | |||             :|| | :
Db       1443 ------VSNLEKKQKKFDQLLAEEKNISARYAEERDRAEAEAREKETKALSLARALEEAL 1496

Qy        486 SLLEKASRQDRELLARLEKELKKVSDVA------GETQGSLSVAQDELVTFSEEL----- 534
                |:| ||:::| | :|  :    ||        ::: :|    :|: |  |||
Db       1497 EAREEAERQNKQLRADMEDLMSSKDDVGKNVHELEKSKRALEQQVEEMRTQLEELEDELQ 1556

Qy        535 ----ANLYHHVCMC---------------NNETPNRVMLDYYREGQGGAGRTSPGGRTSP 575
                  | |   | |                 ||   |:::   ||
Db       1557 ATEDAKLRLEVNMQAMKAQFERDLQTRDEQNEEKKRLLIKQVRE---------------- 1600

Qy        576 EARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRRE-----------P 624
                                                 | : | | |::
Db       1601 -----------------------------------LEAELEDERKQRALAVASKKKME 1623

Qy        625 MNIYNLIAII------RD----QIKHLQAAV-DRTTELSRQRIASQEL---GPAVDKDKE 670
               ::: :| | |      ||    |:: ||| : |   ||   | :  |:       :|  :
Db       1624 IDLKDLEAQIEAANKARDEVIKQLRKLQAQMKDYQRELEEARASRDEIFAQSKESEKKLK 1683

Qy        671 ALMEEILKLKSLLST----------KREQI-------TTLRTVLKANKQTAEVALANLKS 713
               :|  |||:|:  |::          :|:::        : :: |   |:  |  :| |:
Db       1684 SLEAEILQLQEELASSERARRHAEQERDELADEIANSASGKSALLDEKRRLEARIAQLEE 1743

Qy        714 KYENEKA---MVTETMMKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAA 770
               : | |::   :: :   |   ::  |  : |   |      | : |    ||:   ::| |
Db       1744 ELEEEQSNMELLNDRFRKTTLQVDTLNTELAAERSA----AQKSDNARQQLERQNKELKA 1799

Qy        771 AEDEKKTLNSLLRMAIQQKL-----ALTQRLELLELDHEQTRRGRAKA 813
                    |       |  |:: |      ||  ::  ||   ||  : || |
Db       1800 KLQE-------LEGAVKSKFKATISALEAKIGQLEEQLEQEAKERAAA 1840


RESULT 10
MYHB_CHICK
ID   MYHB_CHICK     STANDARD;      PRT;  1978 AA.
AC   P10587;
DT   01-JUL-1989 (Rel. 11, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, gizzard smooth muscle.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88118918; PubMed=2892941;
RA   Yanagisawa M., Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
RA   Masaki T.;
RT   "Complete primary structure of vertebrate smooth muscle myosin heavy
RT   chain deduced from its complementary DNA sequence. Implications on
RT   topography and function of myosin.";
RL   J. Mol. Biol. 198:143-157(1987).
RN   [2]
RP   REVISIONS.
RA   Masaki T.;
RL   Submitted (FEB-1989) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-203.
RX   MEDLINE=88032919; PubMed=3312184;
RA   Maita T., Onishi H., Yajima E., Matsuda G.;
RT   "Amino acid sequence of the amino-terminal 24 kDa fragment of the
RT   heavy chain of chicken gizzard myosin.";
RL   J. Biochem. 102:133-145(1987).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
RX   MEDLINE=98412652; PubMed=9741621;
RA   Dominguez R., Freyzon Y., Trybus K.M., Cohen C.;
RT   "Crystal structure of a vertebrate smooth muscle myosin motor domain
RT   and its complex with the essential light chain: visualization of the
RT   pre-power stroke state.";
RL   Cell 94:559-571(1998).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06546; CAA29793.1; -.
```

```
DR   PIR; S03166; S03166.
DR   PDB; 1BR1; 09-SEP-98.
DR   PDB; 1BR2; 09-SEP-98.
DR   PDB; 1BR4; 09-SEP-98.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Methylation; Alkylation; Calmodulin-binding;
KW   Multigene family; 3D-structure.
FT   INIT_MET      0      0
FT   DOMAIN        1    790       MYOSIN HEAD-LIKE.
FT   DOMAIN      791    820       IQ.
FT   DOMAIN      849   1978       RODLIKE TAIL (S2 AND LMM DOMAINS).
FT   DOMAIN      849   1978       COILED COIL (POTENTIAL).
FT   NP_BIND     176    183       ATP.
FT   DOMAIN      666    688       ACTIN-BINDING.
FT   DOMAIN      767    781       ACTIN-BINDING.
FT   MOD_RES       1      1       BLOCKED.
FT   MOD_RES     127    127       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     706    706       ALKYLATION (SH-1).
FT   MOD_RES     716    716       ALKYLATION (SH-2).
FT   CONFLICT    127    127       MISSING (IN REF. 3).
FT   CONFLICT    204    215       KDTSITQGPSFS -> RTPASLKVHLFP (IN REF. 1).
SQ   SEQUENCE   1978 AA;  228663 MW;  B7B6C923E5273D93 CRC64;

  Query Match          8.3%;  Score 338.5;  DB 1;  Length 1978;
  Best Local Similarity   20.4%;  Pred. No. 1.6e-06;
  Matches  196;  Conservative  147;  Mismatches  335;  Indels  281;  Gaps   31;

Qy         20 EWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEE--------KHQLKLQFEELEVDYE 69
              | |:| : ||  | ::     :||:  |  :: |:|        ::: : |  :|  : |
Db       1095 EELQAALARLEDETSQKNNALKKIRELESHISDLQEDLESEKAARNKAEKQKRDLSEELE 1154

Qy         70 AIRSEME----------QLKEAFGQAHTNHKKVAADGESREESLIQESASKEQYYVRKVL 119
              |:::|:|          :|:    |  |  |:   :     |: :||   |   : : |
Db       1155 ALKTELEDTLDTTATQQELRAKREQEVTVLKRALEEETRTHEAQVQEMRQK---HTQAVE 1211

Qy        120 ELQTELKQLRNVLTNTQSENERL----ASVAQELKEINQ----------NVEIQ------ 159
              ||  :|:|  :   |    : |    | :| |:: ::|          :|:|
Db       1212 ELTEQLEQFKRAKANLDKTKQTLEKDNADLANEIRSLSQAKQDVEHKKKKLEVQLQDLQS 1271

Qy        160 --------RGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVEFEGLKHE 211
                     |  | : : : :     :    :| | :|| | | |: |     : : |  |
Db       1272 KYSDGERVRTELNEKVHKLQIEVENVTSLLNEAESKNIKLTKDVATLGSQLQDTQELLQE 1331

Qy        212 -----------IKRLEEETEYLNSQLEDAIRLKEISER--------------QLEEALET 246
                        :::||::   |  ||:: :  |:  ||              :|:|   |
Db       1332 ETRQKLNVTTKLRQLEDDKNSLQEQLDEEVEAKQNLERHISTLTIQLSDSKKKLQEFTAT 1391

Qy        247 LKTEREQKNSLRKELSHYMSINDSF----------YTSHLHVSLDGL----------- 283
              ::|  | |  |::|:    |:   |              : |   || |
Db       1392 VETMEEGKKKLQREIE---SLTQQFEEKAASYDKLEKTKNRLQQELDDLVVDLDNQRQLV 1448

Qy        284 --------KFSDDAAEPNN-----------DAEA---------LVNGFEHGGLAKLPLD 314
                     ||    ||  |            :|||         |    |    ||  |:
Db       1449 SNLEKKQKKFDQMLAEEKNISSKYADERDRAEAEAREKETKALSLARALEEALEAKEELE 1508

Qy        315 NKTSTPKKEGLAPPSPSLVSDLLSE-----LNISEIQKLKQQLMQMEREKAGLLATLQDT 369
                 | |          : ||:|       |: |::| |: | |           :::
Db       1509 RTNKMLKAE---------MEDLVSSKDDVGKNVHELEKSKRTLEQQ----------VEEM 1549

Qy        370 QKQLEHTRGSLSEQQEKVTRLTENLSAL-----RRLQASKERQTALDNEKDRDSHEDGDY 424
              : |||      |   ::   ||  |: |:     | |||  |:     ||: |
Db       1550 KTQLEELEDELQAAEDAKLRLEVNMQAMKSQFERDLQARDEQ-----NEEKR-------- 1596

Qy        425 YEVDINGPEILACKYHVAVAEAGELREQLKALRSTHEA-------REAQHAEEKGRYEAE 477
                                          |: || |   ||         : |  |  | : | :
Db       1597 -------------------------RQLLKQLHE-HETELEDERKQRALAAAAKKKLEVD 1630

Qy        478 GQALTEKVSLLEKASRQDRELLARLEKELK--------------KVSDVAGETQGSLSVA 523
              : |  :|    ||  :  : | :|: ::|               ::   | | :
Db       1631 VKDLESQVDSANKAREEAIKQLRKLQAQMKDYQRDLDDARAAREEIFATARENEKKAKNL 1690

Qy        524 QDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSP 583
              : ||:   |:||                 |   |  :|       ::  ||||  :   ||
Db       1691 EAELIQLQEDLAAAER----------ARKQADLEKEEMAEELASANSGRTSLQDEKRR-- 1738

Qy        584 ILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIRDQI-KHLQ 642
                                        | : ::    |    :: |  : |:: | :|
Db       1739 ----------------------------LEARIAQLEEELDEEHSNIETMSDRMRKAVQ 1769

Qy        643 AAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLRTVLKANKQ 702
                |     ||: :|  :|       |:: |  :   : | | |  :|    :::  |:
Db       1770 QAEQLNNELATERATAQ-------KNENARQQLERQNKELRSKLQEMEGAVKSKFKSTIA 1822

Qy        703 TAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLR---AMFATRCDEYIT 759
                |  :|:|: : | |          || : | ||:       |     :  : ::
Db       1823 ALEAKIASLEEQLEQEAREKQAAAKTLRQKDKKLKDALLQVEDERKQAEQYKDQAEKGNL 1882

Qy        760 QLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRAKAAPKTK 818
              :| :::|||  ||:| : :|     | ::||   || || |         ||  || |:|
Db       1883 RLKQLKRQLEEAEEESQRIN-----ANRRKL---QR-ELDEATESNDALGREVAALKSK 1932


RESULT 11
MYH9_RAT
ID   MYH9_RAT       STANDARD;      PRT;  1961 AA.
AC   Q62812;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE   type A) (Nonmuscle myosin heavy chain-A) (NMMHC-A).
GN   MYH9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING.
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U31463; AAA74950.1; -.
DR   HSSP; P10587; 1BR2.
```

```
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Alkylation; Multigene family.
FT   DOMAIN        1    778       MYOSIN HEAD-LIKE.
FT   DOMAIN      779    808       IQ.
FT   DOMAIN      841   1927       COILED COIL (POTENTIAL).
FT   NP_BIND     174    181       ATP (POTENTIAL).
FT   DOMAIN      654    676       ACTIN-BINDING.
FT   MOD_RES     694    694       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     704    704       ALKYLATION (SH-2) (POTENTIAL).
SQ   SEQUENCE   1961 AA;  226336 MW;  9B9876D9681FB19E CRC64;

  Query Match             8.2%;  Score 337.5;  DB 1;  Length 1961;
  Best Local Similarity   22.6%;  Pred. No. 1.7e-06;
  Matches  208;  Conservative  175;  Mismatches  358;  Indels  179;  Gaps   39;

Qy          7 EEEYARLVME--AQPEWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEE--------K 54
              : : | | |:   : | |:| : |:  | |:      :||:  |  ::.|:|        :
Db       1068 QAQIAELKMQLAKKEEELQAALARVEEEAAQKNMALKKIRELETQISELQEDLESERACR 1127

Qy         55 HQLKLQFEELEVDYEAIRSEME----------QLKEAFGQAHTNHKKVAADGESREESLI 104
              :: : |  :|  : ||:::|:|          :|:    |  :  ||   |     |: |
Db       1128 NKAEKQKRDLGEELEALKTELEDTLDSTAAQQELRSKREQEVSILKKTLEDEAKTHEAQI 1187

Qy        105 QESASKEQYYVRKVLELQTELKQLRNVLTNTQS--ENERLASVAQELKEINQ---NVEIQ 159
              ||   |    | :: |.  : |:::  |   :   ||||   :| |:| : |   : | :
Db       1188 QEMRQKHSQAVEELAEQLEQTKRVKATLEKAKQTLENER-GELANEVKALLQGKGDSEHK 1246

Qy        160 RGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVEFE---GLKHE----I 212
              | ::   ::| : :       :|| |     |  :|| |   ||| :   || ::
Db       1247 RKKVEAQLQELQVK-------FSEGERVRTELADKVSKL---QVELDSVTGLLNQSDSKS 1296

Qy        213 KRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSH-----YMSI 267
              :| ::   | |||:|   | :   ||       ||   ::||| |::|          ::
Db       1297 SKLTKDFSALESQLQDTQELLQEENRQKLSLSTKLKQMEDEKNSFREQLEEEEEEAKRNL 1356

Qy        268 NDSFYTSHLHVS------LDGLKFSDDAAEP----NNDAEALVNGFEHGGLAKLPLDNKT 317
                | |  |:       ||:   : |.|       | | |     |    |   |: ||
Db       1357 EKQIATLHAQVTDMKKKMEDGVGCLETAEEAKRRLQKDLEGLSQRLEEKVAAYDKLE-KT 1415

Qy        318 STPKKEGLAPPSPSLVSDLLSEL-----NISEIQKLKQQLMQMEREKAGLLATLQDTQKQ 372
              |  :: |          ||| :|     ::| ::| :::  |:  |:  : |   : : :
Db       1416 KTRLQQEL--------DDLLVDLDHQRQSVSNLEKKQKKFDQLLAEEKTISAKYAEERDR 1467

Qy        373 LEHTRGSLSEQQEKVTRLTENLSALRRLQASKERQTALD--NEKDRDSHEDGDYYEVDIN 430
              |.      :| :|| |:    ||  | |: : |::  |:  |:: |   ||    : |:
Db       1468 AE------AEAREKETKA---LSLARALEEAMEQKAELERLNKQFRTEMEDLMSSKDDV- 1517

Qy        431 GPEI--LACKYHVAVAEAGELREQLK----ALRSTHEAR---EAQHAEEKGRYEAEGQAL 481
              |  : |          :  |:: ||:     |::| :|:   |       | ::| : |
Db       1518 GKSVHELEKSNRALEQQVEEMKTQLEELEDELQATEDAKLRLEVNLQAMKAQFERDLQGR 1577

Qy        482 TEKVSLLEK-ASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHH 540
              |:    :|   || ||: | || | |          | |:::|  : :    :| :|  |
Db       1578 DEQSEEKKKQLVRQVREMEAELEDERK---------QRSIAMAARKKLEM--DLKDLEAH 1626

Qy        541 VCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPEARGRRSPIL--------------- 585
              :     :| |:   :  :: :    :      |   : |  |  ||
Db       1627 I-----DTANKNREEAIKQLRKLQAQMKDCMRDVDDTRASREEILAQAKENEKKLKSMEA 1681

Qy        586 ----LPKGLLAPEAGR------ADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIA--- 632
                  | : | | |  :       |    : : | |    : | : ||    :  |||
Db       1682 EMIQLQEELAAAERAKRQAQQERDELADEIANSSGKGALA-LEEKRR----LEALIALLE 1736

Qy        633 -----------IIRDQIKHLQAAVDR-TTELSRQRIASQELGPAVDKDKEALMEEILKLK 680
                        :| |::|     :|:  |:|: :|  :|       |::  |  :   : |
Db       1737 EELEEEQGNTELINDRLKKANLQIDQINTDLNLERSHAQ-------KNENARQQLERQNK 1789

Qy        681 SLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDA 740
              | :  :|  : :::  ||:    |  :| |: : :||         ::|   | ||:
Db       1790 ELKAKLQEMESAVKSKYKASIAALEAKIAQLEEQLDNETKERQAASKQVRRAEKKLKDVL 1849

Qy        741 ATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKL-------. 790
                  |     |  : |:  |:| :::|||  ||:| :  |:  |  :|::|
Db       1850 LQVEDERRNAEQFKDQADKASTRLKQLKRQLEEAEEEAQRANA-SRRKLQRELEDATETA 1908

Qy        791 -ALTQRLELLELDHEQTRRG 809
               |: : :  |:    : |||
Db       1909 DAMNREVSSLK---NKLRRG 1925


RESULT 12
MYHB_MOUSE
ID   MYHB_MOUSE     STANDARD;      PRT;  1972 AA.
AC   O08638; O08639; Q62462; Q64195;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, smooth muscle isoform (SMMHC).
GN   MYH11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Uterus;
RX   MEDLINE=97242182; PubMed=9125171;
RA   Hasegawa K., Arakawa E., Oda S., Matsuda Y.;
RT   "Molecular cloning and expression of murine smooth muscle myosin heavy
RT   chains.";
RL   Biochem. Biophys. Res. Commun. 232:313-316(1997).
RN   [2]
RP   SEQUENCE OF 1-368 FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=95008063; PubMed=7923625;
RA   Miano J.M., Cserjesi P., Ligon K.L., Periasamy M., Olson E.N.;
RT   "Smooth muscle myosin heavy chain exclusively marks the smooth muscle
RT   lineage during mouse embryogenesis.";
RL   Circ. Res. 75:803-812(1994).
RN   [3]
RP   SEQUENCE OF 1-126 FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=96172919; PubMed=8593698;
RA   Suzuki T., Kim H.S., Kurabayashi M., Hamada H., Fujii H., Aikawa M.,
RA   Watanabe M., Watanabe N., Sakomura Y., Yazaki Y., Nagai R.;
RT   "Preferential differentiation of P19 mouse embryonal carcinoma cells
RT   into smooth muscle cells. Use of retinoic acid and antisense against
RT   the central nervous system-specific POU transcription factor Brn-2.";
RL   Circ. Res. 78:395-404(1996).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
```

```
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D85923; BAA19690.1; -.
DR   EMBL; D85924; BAA19691.1; -.
DR   EMBL; L25860; AAA67552.1; -.
DR   EMBL; S81516; AAB36168.1; -.
DR   HSSP; P10587; 1BR2.
DR   MGD; MGI:102643; Myh11.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW   Multigene family; Alternative splicing.
FT   DOMAIN       1    785       MYOSIN HEAD-LIKE.
FT   DOMAIN     786    815       IQ.
FT   DOMAIN     844   1934       COILED COIL (POTENTIAL).
FT   DOMAIN    1935   1972       CARBOXYL-TERMINAL.
FT   NP_BIND    178    185       ATP (POTENTIAL).
FT   DOMAIN     661    683       ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN     762    776       ACTIN-BINDING (BY SIMILARITY).
FT   MOD_RES    129    129       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES    701    701       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES    711    711       ALKYLATION (SH-2) (POTENTIAL).
FT   VARSPLIC  1930   1972       RGNEASFVPSRRAGGRRVIENTDGSEEEMDARDSDFNGTKA
FT                               SE -> GPPPQETSQ (IN ISOFORM 2).
FT   CONFLICT   126    126       N -> D (IN REF. 3).
FT   CONFLICT   161    161       A -> V (IN REF. 2).
FT   CONFLICT   189    189       Q -> K (IN REF. 2).
SQ   SEQUENCE  1972 AA;  227026 MW;  A1398E3F5B11F15A CRC64;

  Query Match             8.2%;  Score 335.5;  DB 1;  Length 1972;
  Best Local Similarity   23.3%;  Pred. No. 2.1e-06;
  Matches  211;  Conservative  165;  Mismatches  382;  Indels  149;  Gaps   38;

Qy          7 EEEYARLVME--AQPEWLRAEVKRLSHELAETTR--EKIQAAEYGLAVLEE--------K 54
              : : | | |:   : | |:| : ||  |:|:     :||:  |  :: |:|        :
Db       1075 QAQIAELKMQLAKKEEELQAALARLDEEIAQKNNALKKIRELEGHISDLQEDLDSERAAR 1134

Qy         55 HQLKLQFEELEVDYEAIRSEME----------QLKEAFGQAHTNHKKVAADGESR-EESL 103
              :: : |  :|  : ||:::|:|          :|:    |  |  || | | |:|  |:
Db       1135 NKAEKQKRDLGEELEALKTELEDTLDSTATQQELRAKREQEVTVLKK-ALDEETRSHEAQ 1193

Qy        104 IQESASKEQYYVRKVLELQTELKQLRNVLTN-------TQSENERLASVAQELKEINQNV 156
              :||   |   : : | ||  :|:| :    |        : ||  ||   : | :  | |
Db       1194 VQEMRQK---HTQAVEELTEQLEQFKRAKANLDKSKQTLEKENADLAGELRVLGQAKQEV 1250

Qy        157 EIQRGRLRDDIKEYKFREARLLQDYSELEEENISLQKQVSVLRQNQVE-FEGLKHEIK-- 213
              | :: :|   :::       |    |: |     |  :|  | ||:||   |: :| :
Db       1251 EHKKKKLEVQLQD-------LQSKCSDGERARAELSDKVHKL-QNEVESVTGMLNEAEGK 1302

Qy        214 --RLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKNSLRKELSHYMSINDSF 271
                :| ::   | |||:|   | :   ||       |:   :::|||: :|   |    :
Db       1303 AIKLAKDVASLGSQLQDTQELLQEETRQKLNVSTKLRQLEDERNSLQDQLDEEMEAKQNL 1362

Qy        272 YTSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHG---------GLAKLPLDNKTSTPKK 322
                |||   :: | |: :   |  : :   | |          ||::   :  :  |
Db       1363 ---ERHVSTLNIQLS-DSKKKLQDFASTIEVMEEGKKRLQKEMEGLSQQYEEKAAAYDKL 1418

Qy        323 EGLAPPSPSLVSDLLSELN-----ISEIQKLKQQLMQMEREKAGLLATLQDTQKQLEHTR 377
              |         : ||: :|:     :| ::| ::: |:  |:  : :   | : : |
Db       1419 EKTKNRLQQELDDLVVDLDNQRQLVSNLEKKQKKFDQLLAEEKNISSKYADERDRAE--- 1475

Qy        378 GSLSEQQEKVTRLTENLSALRRLQASKERQTALDNE--------KDRDSHEDGDYYEVDI 429
               :| :|| |:    ||  | |: : | :  |:          :|  | :|    :|
Db       1476 ---AEAREKETKA---LSLARALEEALEAKEELERTNKMLKAEMEDLVSSKD----DVGK 1525

Qy        430 NGPEILACKYHVAVAEAGELREQLKA----LRSTHEAR---EAQHAEEKGRYEAEGQALT 482
              |  |:  |       :  |:: ||:     :::| :|:   |      ||::| : ||
Db       1526 NVHELEKSK-RALETQMEEMKTQLEESEDDVQATEDAKLRLEVNMQALKGQFERDLQARD 1584

Qy        483 EK-VSLLEKASRQDRELLARLEKELKKVSDVAG---ETQG---SLSVAQDELVTFSEELA 535
              |:      :  ||  |    || | |: :  |   : :|    | :  |  :   ||
Db       1585 EQNEEKRRQLQRQLHEYETELEDERKQRALAAAAKKKLEGDLKDLELQADSAIKGREEAI 1644

Qy        536 NLYHHVCMCNNETPNRVMLDYYREGQGGAGRTSPGGRTSPE----ARGRRSPIL-LPKGL 590
                :           | |: ||              || |    |:   : :: | : |
Db       1645 KQLRKL--------QAQMKDFQRELDDARASRDEIFATSKENEKKAKSLEADLMQLQEDL 1696

Qy        591 LAPEAGR--ADGGTGDSSPSPGSSLP--SPLSD-PRREPMNIYNLIAIIRDQIKHLQAAV 645
              | |  | ||    : :   |||   : | |  ||    | |   : ::  :::|
Db       1697 AAAERARKQADLEKEELAEELASSLSGRNTLQDEKRRLEARIAQLEEELEEEQGNMEAMS 1756

Qy        646 DR-----------TTELSRQRIASQELGPAVDKDKEALMEEILKLKSLLSTKREQITTLR 694
              ||           : ||: :|  :|       |:: |  :   :  | | |  :|    ::
Db       1757 DRVRKATLQAEQLSNELATERSTAQ-------KNESARQQLERQNKELRSKLQEVEGAVK 1809

Qy        695 TVLKANKQTAEVALANLKSKYENEKAMVTETMMKLRNELKALKEDAATFSSLRAM---FA 751
              ||:     |  :| |: : | |          |: : | |||        | |   :
Db       1810 AKLKSTVAALEAKIAQLEEQVEQEAREKQAATKSLKQKDKKLKEVLLQVEDERKMAEQYK 1869

Qy        752 TRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQKLALTQRLELLELDHEQTRRGRA 811
              : ::  |:: :::||| ||:| : :|     |.::||   || || |        ||
Db       1870 EQAEKGNTKVKQLKRQLEEAEEESQCIN-----ANRRKL---QR-ELDEATESNEAMGRE 1920

Qy        812 KAAPKTK 818
                | |:|
Db       1921 VNALKSK 1927


RESULT 13
MYH9_HUMAN
ID   MYH9_HUMAN     STANDARD;      PRT;  1960 AA.
AC   P35579; O60805;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE   type A) (Nonmuscle myosin heavy chain-A) (NMMHC-A).
GN   MYH9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
```

```
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA   Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA   Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA   McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA   Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA   Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA   Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA   Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [2]
RP   SEQUENCE OF 1-1337 FROM N.A.
RX   MEDLINE=92003925; PubMed=1912569;
RA   Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA   Arnaout M.A., Clayton L.K., Tenen D.G.;
RT   "Cellular myosin heavy chain in human leukocytes: isolation of 5'
RT   cDNA clones, characterization of the protein, chromosomal
RT   localization, and upregulation during myeloid differentiation.";
RL   Blood 78:1826-1833(1991).
RN   [3]
RP   SEQUENCE OF 1-715 FROM N.A.
RX   MEDLINE=91316803; PubMed=1860190;
RA   Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA   Gdula D., Adelstein R.S., Weir L.;
RT   "Human nonmuscle myosin heavy chains are encoded by two genes located
RT   on different chromosomes.";
RL   Circ. Res. 69:530-539(1991).
RN   [4]
RP   SEQUENCE OF 714-1960 FROM N.A.
RX   MEDLINE=90138958; PubMed=1967836;
RA   Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
RT   "Human nonmuscle myosin heavy chain mRNA: generation of diversity
RT   through alternative polyadenylylation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
RN   [5]
RP   VARIANT DFNA17 HIS-705.
RX   MEDLINE=20489856; PubMed=11023810;
RA   Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
RA   Mhatre A.N.;
RT   "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
RT   nonmuscle myosin MYH9.";
RL   Am. J. Hum. Genet. 67:1121-1128(2000).
RN   [6]
RP   VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
RX   MEDLINE=20428192; PubMed=10973259;
RA   Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
RA   Ghiggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
RA   Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
RA   Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,
RA   Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
RT   "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
RT   Sebastian syndromes.";
RL   Nat. Genet. 26:103-105(2000).
RN   [7]
RP   VARIANTS MHA ILE-1155 AND LYS-1841.
RX   MEDLINE=20428193; PubMed=10973260;
RA   Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
RT   "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
RT   May-Hegglin anomaly.";
RL   Nat. Genet. 26:106-108(2000).
CC   -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC       CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC       CAPPING.
CC   -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC       CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC       REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
CC       (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
CC       (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC       WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL
CC       DEAFNESS, CATARACTS AND NEPHRITIS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
CC       (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
CC       BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
CC   -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
CC       FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
CC       CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
CC       COCHLEOSACCULAR DEGENERATION.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z82215; CAB05105.1; -.
DR   EMBL; M81105; AAA59888.1; -.
DR   EMBL; M69180; AAA61765.1; -.
DR   EMBL; M31013; AAA36349.1; -.
DR   HSSP; P10587; 1BR2.
DR   Genew; HGNC:7579; MYH9.
DR   MIM; 160775; -.
DR   MIM; 153640; -.
DR   MIM; 155100; -.
DR   MIM; 603622; -.
DR   MIM; 605249; -.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
```

```
KW   Coiled coil; Alkylation; Multigene family; Disease mutation;
KW   Deafness.
FT   DOMAIN        1    778       MYOSIN HEAD-LIKE.
FT   DOMAIN      779    808       IQ.
FT   DOMAIN      837   1926       COILED COIL (POTENTIAL).
FT   NP_BIND     174    181       ATP (POTENTIAL).
FT   DOMAIN      654    676       ACTIN-BINDING.
FT   MOD_RES     694    694       ALKYLATION (SH-1) (POTENTIAL).
FT   MOD_RES     704    704       ALKYLATION (SH-2) (POTENTIAL).
FT   VARIANT      93     93       N -> K (IN MHA).
FT                                /FTId=VAR_010791.
FT   VARIANT     702    702       R -> C (IN FTNS).
FT                                /FTId=VAR_010792.
FT   VARIANT     705    705       R -> H (IN DFNA17).
FT                                /FTId=VAR_010793.
FT   VARIANT    1155   1155       T -> I (IN MHA).
FT                                /FTId=VAR_010794.
FT   VARIANT    1165   1165       R -> C (IN SBS).
FT                                /FTId=VAR_010795.
FT   VARIANT    1424   1424       D -> H (IN FTNS).
FT                                /FTId=VAR_010796.
FT   VARIANT    1841   1841       E -> K (IN MHA).
FT                                /FTId=VAR_010797.
FT   CONFLICT     53     55       EAI -> RGH (IN REF. 3).
FT   CONFLICT    660    660       T -> S (IN REF. 3).
FT   CONFLICT    869    869       T -> M (IN REF. 4).
FT   CONFLICT    931    931       C -> Y (IN REF. 4).
FT   CONFLICT   1240   1241       KG -> GR (IN REF. 4).
FT   CONFLICT   1350   1350       E -> EE (IN REF. 2).
FT   CONFLICT   1764   1764       T -> A (IN REF. 2).
FT   CONFLICT   1771   1771       S -> G (IN REF. 2).

  Query Match              8.2%;  Score 335;  DB 1;  Length 1960;
  Best Local Similarity   20.3%;  Pred. No. 2.1e-06;
  Matches  214;  Conservative  159;  Mismatches  351;  Indels  328;  Gaps   37;

Qy          6 EEEEYAR---------------------LVMEAQPEWLRAEVKRLSHELAETTREKIQAA 44
                |||| ||                     :::| |   |  | | |   :|| |     :
Db        953 EEEESARQKLQLEKVTTEAKLKKLEEEQIILEDQNCKLAKEKKLLEDRIAEFTTNLTEEE 1012

Qy         45 E--YGLAVLEEKHQLKL------------QFEELEVDYEAIRSEMEQLKEAFGQAHTNHK 90
              |    || |: ||:  :            | :|||       :  :   |  :    :
Db       1013 EKSKSLAKLKNKHEAMITDLEERLRREEKQRQELEKTRRKLEGDSTDLSDQIAELQAQIA 1072

Qy         91 KVAADGESREESL------IQESASKEQYYVRKVLELQTELKQLRNVLTNTQSENERLAS 144
              ::        :|| |      ::| |:::    ::|: ||::::  :|:   |    :||
Db       1073 ELKMQLAKKEEELQAALARVEEEAAQKNMALKKIRELESQISELQEDL---ESERASRNK 1129

Qy        145 VAQELKEINQNVEIQRGRLRD---------DIKEYKFREARLLQDYSELEEENISLQKQV 195
                :: ::: : :|  :   | |          :::   : :|  :|:    |||  : : |:
Db       1130 AEKQKRDLGEELEALKTELEDTLDSTAAQQELRSKREQEVNILK--KTLEEEAKTHEAQI 1187

Qy        196 SVLRQNQVEFEGLKHEIKRLEEETEYLNSQLEDAIRLKEISERQLEEALETLKTEREQKN 255
                :||          ||        :   |  |  |||    |:|     ||:|  :||: || :
Db       1188 QEMRQ--------KH-----SQAVEELAEQLEQTKRVK----ANLEKAKQTLENERGE-- 1228

Qy        256 SLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNNDAEA----LVNGFEHGGLAKL 311
                           : : |  | |        |:         ||       |    |  |   :
Db       1229 ----------------LANEVKVLLQG---KGDSEHKRKKVEAQLQELQVKFNEGERVRT 1269

Qy        312 PLDNKTSTPKKE-----GLAPPSPSLVSDLLSELNISEIQ------------------- 346
               | :|  :  : |        ||    |  |   |  |  :  :   | |
Db       1270 ELADKVTKLQVELDNVTGLLSQSDSKSSKLTKDFSALESQLQDTQELLQEENRQKLSLST 1329

Qy        347 KLKQ----------QLMQ-------MEREKAGLLATLQDTQKQLEHTRGSLSEQQEKVTR 389
              ||||            || :        :|:: | | |  : | :|::| : | |    :|   :
Db       1330 KLKQVEDEKNSFREQLEEEEEAKHNLEKQIATLHAQVADMKKKMEDSVGCLETAEEVKRK 1389

Qy        390 LTENLSALR-----------RLQASK----------------ERQTALDNEKDRDSH--- 419
               | ::| |                 :|: :|                    :||:| : || :
Db       1390 LQKDLEGLSQRHEEKVAAYDKLEKTKTRLQQELDDLLVDLDHQRQSACNLEKKQKKFDQL 1449

Qy        420 ------------EDGDYYEVDINGPEILACKYHVAVAEAGELREQLKAL----------- 456
                          |: |  | :      |  |      |: || | : :|: |
Db       1450 LAEEKTISAKYAEERDRAEAEAREKETKALSLARALEEAMEQKAELERLNKQFRTEMEDL 1509

Qy        457 --------RSTHEAREAQHAEE-----------------------------------KGRY 474
                       :| ||  ::: | |                                   | ::
Db       1510 MSSKDDVGKSVHELEKSKRALEQQVEEMKTQLEELEDELQATEDAKLRLEVNLQAMKAQF 1569

Qy        475 EAEGQALTEKVSLLEK-ASRQDRELLARLEKELKKVSDVAGETQGSLSVAQDELVTFSEE 533
              | : |     |:     :|   || ||: | || | |          | |::||  : :    :
Db       1570 ERDLQGRDEQSEEKKKQLVRQVREMEAELEDERK---------QRSMAVAARKKLEM--D 1618

Qy        534 LANLYHHVCMCNNETPNRV---------MLDYYREGQGGAGRTSPGGRTSPEARGRRSPI 584
              | :|  |:   |          :               | |  ||                :  |  |  |
Db       1619 LKDLEAHIDSANKNRDEAIKQLRKLQAQMKDCMRE--------------LDDTRASREEI 1664

Qy        585 L-------------------LPKGLLAPEAGR------ADGGTGDSSPSPGSSLPSPLSD 619
              |                   | : | | |  :          |        : : | |     : | :
Db       1665 LAQAKENEKKLKSMEAEMIQLQEELAAAERAKRQAQQERDELADEIANSSGKGALA-LEE 1723

Qy        620 PRREPMNIYNL----------IAIIRDQIKHLQAAVDR-TTELSRQRIASQELGPAVDKD 668
                ||   | |                :| |::|      :|:  |:|: :|  :|              |:
Db       1724 KRRLEARIAQLEEELEEEQGNTELINDRLKKANLQIDQINTDLNLERSHAQ-------KN 1776

Qy        669 KEALMEEILKLKSLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMMK 728
               : |  :    : | |     :|    |:::  ||:   |    :| |: : :||            :
Db       1777 ENARQQLERQNKELKVKLQEMEGTVKSKYKASITALEAKIAQLEEQLDNETKERQAACKQ 1836

Qy        729 LRNELKALKEDAATFSSLR---AMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMA 785
               :|   | ||:            |          :   : |:  :|   |:| :::|||  ||:|  :   |:   |
Db       1837 VRRTEKKLKDVLLQVDDERRNAEQYKDQADKASTRLKQLKRQLEEAEEEAQRANA-SRRK 1895

Qy        786 IQQKL--------ALTQRLELLELDHEQTRRG 809
               :|::|          |: : :· |:       : |||
Db       1896 LQRELEDATETADAMNREVSSLK---NKLRRG 1924


RESULT 14
MYSA_DROME
ID   MYSA_DROME     STANDARD;      PRT;  1962 AA.
AC   P05661;
DT   01-NOV-1988 (Rel. 09, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin heavy chain, muscle.
GN   MHC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=Canton-S; TISSUE=Pupae;
RX   MEDLINE=89384556; PubMed=2506434;
RA   George E.L., Ober M.B., Emerson C.P. Jr.;
RT   "Functional domains of the Drosophila melanogaster muscle myosin
RT   heavy-chain gene are encoded by alternatively spliced exons.";
RL   Mol. Cell. Biol. 9:2957-2974(1989).
RN   [2]
RP   SEQUENCE OF 1-264 FROM N.A.
RX   MEDLINE=87280141; PubMed=3038896;
RA   Wassenberg D.R. II, Kronert W.A., O'Donnell P.T., Bernstein S.I.;
RT   "Analysis of the 5' end of the Drosophila muscle myosin heavy chain
RT   gene. Alternatively spliced transcripts initiate at a single site and
RT   intron locations are conserved compared to myosin genes of other
RT   organisms.";
RL   J. Biol. Chem. 262:10741-10747(1987).
RN   [3]
RP   SEQUENCE OF 333-614 FROM N.A.
RC   STRAIN=Canton-S; TISSUE=Embryonic muscle;
RX   MEDLINE=91330870; PubMed=1907912;
RA   Kronert W.A., Edwards K.A., Roche E.S., Wells L., Bernstein S.I.;
RT   "Muscle-specific accumulation of Drosophila myosin heavy chains: a
```

```
RT   splicing mutation in an alternative exon results in an isoform
RT   substitution.";
RL   EMBO J. 10:2479-2488(1991).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- ALTERNATIVE PRODUCTS: At least 15 isoforms; AAAAA (shown here),
CC       BDBBA, BABDB, 3b, 7b, 7c, 7d, 9b, 9c, 11b, 11c, 11d, 11e, 15b and
CC       18; may be produced by alternative splicing. Exons 3, 7, 9, 11 and
CC       15 are mutually exclusive splicing exons and exon 18 is included
CC       or excluded.
CC   -!- TISSUE SPECIFICITY: Expressed in larval and adult muscles.
CC       Isoforms containing exon 9a are expressed in indirect flight
CC       muscles, exons 9a and 9b are expressed in jump muscles, exons 9b
CC       and 9c are expressed in other larval and adult muscles.
CC   -!- DOMAIN: Alternative splicing exons contribute to the specialized
CC       contractile activities of different muscle types. Exon 3 encodes
CC       the hydrophobic pocket adjacent to the ATP-binding site, exon 9 is
CC       adjacent to the actin-binding domain, exon 11 is involved in
CC       actin-binding, exon 15 in the S2 hinge and exons 18 and 19 the
CC       non-coiled tail region.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61229; AAA28686.1; -.
DR   EMBL; M61229; AAA28687.1; -.
DR   EMBL; J02788; AAA28706.1; -.
DR   EMBL; J02788; AAA28707.1; -.
DR   EMBL; X60196; CAA42752.1; -.
DR   EMBL; X60196; CAA42753.1; -.
DR   EMBL; X60196; CAA42754.1; -.
DR   PIR; A28492; A28492.
DR   PIR; A32491; A32491.
DR   PIR; B32491; B32491.
DR   HSSP; P08799; 1MND.
DR   FlyBase; FBgn0002741; Mhc.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 7.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 4.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Calmodulin-binding; Alternative splicing;
KW   Multigene family.
FT   DOMAIN        1    779       MYOSIN HEAD-LIKE.
FT   DOMAIN      780    809       IQ.
FT   DOMAIN      802   1927       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP (BY SIMILARITY).
FT   VARSPLIC     69    116       VRDIKSEKVEKVNPPKFEKIEDMADMTVLNTPCVLHNLRQR
FT                                YYAKLIY -> TRDLKKDLLQQVNPPKYEKAEDMSNLTYLN
FT                                DASVLHNLRQRYYNKLIY (IN ISOFORM 3B,
FT                                ISOFORM BDBBA AND ISOFORM BABDB).
FT   VARSPLIC    298    332       DICLLTDNIYDYHIVSQGKVTVASIDDAEEFSLTD -> EY
FT                                CLLSNNIYDYRIVSQGKTTIPSVNDGEEWVAVD (IN
FT                                ISOFORM 7B).
FT   VARSPLIC    298    332       DICLLTDNIYDYHIVSQGKVTVASIDDAEEFSLTD -> EM
FT                                VFLGQHIGDYPGICQGKTRIPGVNDGEEFELTD (IN
FT                                ISOFORM 7C).
FT   VARSPLIC    298    332       DICLLTDNIYDYHIVSQGKVTVASIDDAEEFSLTD -> EM
FT                                CFLSDNIYDYYNVSQGKVTVPNMDDGEEFQLAD (IN
FT                                ISOFORM 7D AND ISOFORM BDBBA).
FT   VARSPLIC    469    525       YNGFEQLCINFTNEKLQQFFNHIMFVMEQEEYKKEGINWDF
FT                                IDFGMDLLACIDLIEK -> YNGFEQLCINFTNEKLQQFFN
FT                                HHMFVLEQEEYKREGIDWAFIDFGMDLLACIDLIEK (IN
FT                                ISOFORM 9B, ISOFORM BDBBA AND ISOFORM
FT                                BABDB).
FT   VARSPLIC    469    525       YNGFEQLCINFTNEKLQQFFNHIMFVMEQEEYKKEGINWDF
FT                                IDFGMDLLACIDLIEK -> YNGFEQLCINFTNEKLQQFFN
FT                                HHMFVLEQEEYQREGIEWTFIDFGMDLQLCIDLIEK (IN
FT                                ISOFORM 9C).
FT   VARSPLIC    723    761       YQILNPRGIKDLDCPKKASKVLIESTELNEDLYRLGHTK
FT                                -> YQILNPAGIVGVDDPKKCGSIILESTALDPDMYRIGHT
FT                                K (IN ISOFORM 11B AND ISOFORM BDBBA).
FT   VARSPLIC    723    761       YQILNPRGIKDLDCPKKASKVLIESTELNEDLYRLGHTK
FT                                -> YQILNPKGIKGIEDPKKCTKVLIESTELNDDQYRLGNT
FT                                K (IN ISOFORM 11C).
FT   VARSPLIC    723    761       YQILNPRGIKDLDCPKKASKVLIESTELNEDLYRLGHTK
FT                                -> YKIMCPKLLQGVEKDKKATEIIIKFIDLPEDQYRLGNT
FT                                K (IN ISOFORM 11D AND ISOFORM BABDB).
FT   VARSPLIC    723    761       YQILNPRGIKDLDCPKKASKVLIESTELNEDLYRLGHTK
FT                                -> YMILAPAIMAAEKVAKNAAGKCLEAVGLDPDMYRIGHT
FT                                K (IN ISOFORM 11E).
FT   VARSPLIC   1216   1241       AEHDRQTCHNELNQTRTACDQLGRDK -> AEKEKNEYYGQ
FT                                LNDLRAGVDHITNEK (IN ISOFORM 15B AND
FT                                ISOFORM BABDB).
FT   VARSPLIC   1936   1936       P -> I (IN ISOFORM 18).
FT   VARSPLIC   1937   1962       MISSING (IN ISOFORM 18).
FT   CONFLICT     43     44       EK -> RE (IN REF. 2).
FT   CONFLICT     68     68       E -> K (IN REF. 2).
FT   CONFLICT    215    215       L -> M (IN REF. 2).
SQ   SEQUENCE   1962 AA;  224480 MW;  F5A888932E414F7F CRC64;
```

Query Match 8.1%; Score 331; DB 1; Length 1962;
Best Local Similarity 21.6%; Pred. No. 3.1e-06;
Matches 200; Conservative 145; Mismatches 357; Indels 224; Gaps 34;

```
Qy        27 KRLSHELAETTREKIQAAEYGLAVLEE-KHQLKLQFEELEVDYE---AIRSEMEQLK--- 79
             |::  |  : | |::||||  :  | : | :|:   :|||   |    :| ::|: |
Db       995 KKMQGETNQKTGEELQAAEDKINHLNKVKAKLEQTLDELEDSLEREKKVRGDVEKSKRKV 1054

Qy        80 --------EAFGQAHTNHKKVAADGESREESLIQESASKEQYYV------RKVLELQTEL 125
                    ||       | |::    : ::: |   :|  |   |      |:: |||  :
Db      1055 EGDLKLTQEAVADLERNKKELEQTIQRKDKELSSITAKLEDEQVVVLKHQRQIKELQARI 1114

Qy       126 KQLRNVL-----TNTQSENERLASVAQELKEINQNVEIQRGRLRDDIKEYKFREARLLQD 180
             ::|   :         ::| :| | :|:||:|: : :|   |     |:  | ||| | :
Db      1115 EELEEEVEAERQARAKAEKQR-ADLARELEELGERLEEAGGATSAQIELNKKREAELSKL 1173

Qy       181 YSELEEENISLQKQVSVLRQNQVEFEGLKHE--IKRLEEETEYLNSQLEDAIRLKEISER 238
               :||| ||  :  :: ||:        ||   :  : |: : ||     |   ::
Db      1174 RRDLEEANIQHESTLANLRK--------KHNDAVAEMAEQVDQLNKLKAKAEHDRQTCHN 1225

Qy       239 QLEE---ALETLKTEREQKNSLRKELSHYMSINDSFYTSHLHVSLDGLKFSDDAAEPNND 295
             :| :   | : |  ::  :  : |:| | :         : :   ||              :
Db      1226 ELNQTRTACDQLGRDKAAQEKIAKQLQHTL--------NEVQSKLD------------E 1264

Qy       296 AEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSLVSDLLSEL-----NISEIQKLKQ 350
               :| |:     || ::|                  ||| :|     :|:: |:|
Db      1265 TNRTLNDFD-ASKKKLSIEN----------------SDLLRQLEEAESQVSQLSKIKI 1305

Qy       351 QL-MQME----------REKAGLLATLQDTQKQLEHTRGSLSEQQEKVTRLTENLS---- 395
               |  |:|          ||:| ||   :: :  |:: |  : |: |    |  ||
Db      1306 SLTTQLEDTKRLADEESRERATLLGKFRNLEHDLDNLREQVEEEAEGKADLQRQLSKANA 1365
```

```
Qy        396 ----------------------ALRRLQAS-----------KERQTALDNEKDRDSHEDG 422
                                     | |:|||             ::   |:  | | | |
Db       1366 EAQVWRSKYESDGVARSEELEEAKRKLQARLAEAEETIESLNQKCIGLEKTKQRLSTEVE 1425

Qy        423 D-YYEVD-----INGPEILACKYHVAVAE----AGELREQLKALRSTHEAREAQHAEEKG 472
              |   |||       |  |     :   : |      :|  :| | :          :     ||
Db       1426 DLQLEVDRANAIANAAEKKQKAFDKIIGEWKLKVDDLAAELDASQKECRNYSTELFRLKG 1485

Qy        473 RYEAEGQALTEKV------------SLLEKASRQDRELLARLEKELKKVSDVAGETQGSL 520
               || |||   | |              ||::     |  :  :||  |::     | | :|
Db       1486 AYE-EGQEQLEAVRRENKNLADEVKDLLDQIGEGGRN-IHEIEKARKRLEAEKDELQAAL 1543

Qy        521 SVAQDELVTFSEELANLYHHVCMCNNETPNRVM-----LDYYREGQGGAGRTSPGGRTSP 575
                |:  |      ::       :      |   |:       :  |:    |   |
Db       1544 EEAEAALEQEENKVLRAQLELSQVRQEIDRRIQEKEEEFENTRKNHQRA-LDSMQASLEA 1602

Qy        576 EARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDPRREPMNIYNLIAIIR 635
               ||:|:   : : | | |                    : |   |    :          |  :
Db       1603 EAKGKAEALRMKKKLEADI----------------NELEIALDHANKANAEAQKNIKRYQ 1646

Qy        636 DQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLKSLL------------ 683
                |:| :| |::   |  | |  ::|     ::   ||  |: : ::||
Db       1647 QQLKDIQTALE---EEQRARDDAREQLGISERRANALQNELEESRTLLEQADRGRRQAEQ 1703

Qy        684 ------------STKREQITTLRTVLKANKQTAEVALANLKSKYENEKAMVTETMM---K 728
                         | :   |:  :  |::  ||    |  | :: :| :    :  |:   :
Db       1704 ELADAHEQLNEVSAQNASISAAKRKLESELQTLHSDLDELLNEAKNSEEKAKKAMVDAAR 1763

Qy        729 LRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAIQQ 788
               | :||:| :: | |   ||      :  |   :|||  :    | :  ||       |||:
Db       1764 LADELRAEQDHAQTQEKLRKALEQQIKELQVRLDEAEAN--ALKGGKK--------AIQK 1813

Qy        789 KLALTQRLELL--ELDHEQTRRGRAK 812
                  | ||:  |  ||| || |   |:
Db       1814 ---LEQRVRELENELDGEQRRHADAQ 1836
```

```
RESULT 15
MYSN_DROME
ID   MYSN_DROME     STANDARD;      PRT;  2017 AA.
AC   Q99323;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain, non-muscle (Zipper protein) (Myosin II).
GN   ZIP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=90349606; PubMed=2117279;
RA   Ketchum A.S., Stewart C.T., Stewart M., Kiehart D.P.;
RT   "Complete sequence of the Drosophila nonmuscle myosin heavy-chain
RT   transcript: conserved sequences in the myosin tail and differential
RT   splicing in the 5' untranslated sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC   -!- FUNCTION: NONMUSCLE MYOSIN APPEARS TO BE RESPONSIBLE FOR
CC       CELLULARIZATION. REQUIRED FOR MORPHOGENESIS AND CYTOKINESIS.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35012; AAA28713.1; -.
DR   PIR; A36014; A36014.
DR   PIR; B36014; B36014.
DR   HSSP; P10587; 1BR2.
DR   FlyBase; FBgn0005634; zip.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Alternative splicing; Coiled coil; Actin-binding;
KW   ATP-binding; Calmodulin-binding.
FT   DOMAIN        1    829       MYOSIN HEAD-LIKE.
FT   DOMAIN      830    859       IQ.
FT   DOMAIN      886   2017       COILED COIL (POTENTIAL).
FT   NP_BIND     225    232       ATP.
FT   DOMAIN      250    260       25 KDA/50 KDA JUNCTION.
FT   DOMAIN      682    694       50 KDA/20 KDA JUNCTION.
FT   DOMAIN      705    727       ACTIN-BINDING.
FT   DOMAIN      742    758       REACTIVE SULFHYDRYL/ACTIN-BINDING.
FT   DOMAIN     1303   2017       LIGHT MEROMYOSIN (LMM).
FT   DOMAIN     1303   1970       ALPHA-HELICAL TAILPIECE (LMM).
FT   DOMAIN     1971   2017       GLOBULAR TAILPIECE.
FT   VARSPLIC      1     45       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   2017 AA;  232016 MW;  73E3CB02BA8F2528 CRC64;
```

```
  Query Match             8.1%;  Score 331;  DB 1;  Length 2017;
  Best Local Similarity   22.5%;  Pred. No. 3.2e-06;
  Matches  198;  Conservative  145;  Mismatches  306;  Indels  230;  Gaps   37;

Qy          2 SAPSEEEEYARLVMEAQPEWLRAEVKRLSHELAETTREKIQAAEYGLAVLEEKHQLKLQF 61
              |  |  :|  |   :|:     :::  |  :|||  | : :  |    : :|   :  |
Db       1286 SVNSSRQENDRRRKQAE-----SQIAELQVKLAEIERARSELQEKCTKLQQEAENITNQL 1340

Qy         62 EELEVDYEAI---RSEME-QLKEA--FGQAHTNHK--------KVAADGESREESLIQES 107
              || |:   |     | || || ||    :  |  |         :: :: |: :| | ::
Db       1341 EEAELKASAAVKSASNMESQLTEAQQLLEEETRQKLGLSSKLRQIESEKEALQEQLEEDD 1400

Qy        108 ASKEQYYVRKVLELQTELKQLRNVLTNTQSENERLASVAQELKE----INQNVEIQRGRL 163
               :|   | ||: |: |::::::       : |  | :|:||:|    :|:::|     |
Db       1401 EAKRN-YERKLAEVTTQMQEIK-------KKAEEDADLAKELEEGKKRLNKDIE----AL 1448

Qy        164 RDDIKEYKFREARLLQD----YSELEEENISLQKQ---VSVLRQNQVEFEGLKHEIKRLE 216
                :||   :  || :      ||||:  | |: |   |  | : |  |: :  | | :
Db       1449 ERQVKELIAQNDRLDKSKKKIQSELEDATIELEAQRTKVLELEKKQKNFDKILAEEKAIS 1508

Qy        217 EETEYLNSQLEDAIRLKEIS----ERQLEEALETLKTEREQKNSLRKELSHYMSINDSFY 272
              |:         |   | ||       |:|:|| : ::    :: :|: ||
Db       1509 EQIAQERDTAEREAREKETKVLSVSRELDEAFDKIEDLENKRKTLQNEL---------- 1557

Qy        273 TSHLHVSLDGLKFSDDAAEPNNDAEALVNGFEHGGLAKLPLDNKTSTPKKEGLAPPSPSL 332
                           || |     |:  |:  |    ||  |::: :  | :
Db       1558 -------------DDLANTQGTADKNVHELEK---AKRALESQLAELKAQN-------- 1592

Qy        333 VSDLLSELNISEIQKLKQQL------MQMEREKAGLLATLQDTQKQLEHTRGSLSEQQEK 386
               :|  :| ::|  ||: ::       | ||:   |||  :  :   |  || : : ::
Db       1593 -EELEDDLQLTEDAKLRLEVNMQALRSQFERD---LLAKEEGAE---EKRRGLVKQLRDL 1645

Qy        387 VTRLTENLSALRRLQASKERQTALDNEKDRDSHEDGDYYEVDINGPEILACKYHVAVAE- 445
              | | |             |:|  |: ::|      :||  |::         : |  | |
Db       1646 ETELDEE---------RKQRTAAVASKKKL----EGDLKEIE------TTMEMHNKVKED 1686

Qy        446 ----AGELREQLK-ALRSTHEAREAQHAEEKGRYEAEGQALTEKVSLLEKASRQDRELLA 500
```

```
                 | :|: |:| |||   ||: |:   :    ||:|     ||  ||    |  | ||
Db       1687 ALKHAKKLQAQVKDALRDAEEAKAAKEELQALSKEADG-----KVKALEAEVLQLTEDLA 1741

Qy        501 RLEKELKKVSDVAGETQGSLSVAQDELVTFSEELANLYHHVCMCNNETPNRVMLDYYREG 560
              |:  :     | ||:      :|||   :||:|         ||     :|:|  |
Db       1742 SSERARR-----AAETE------RDEL---AEEIA---------NNANKGSLMIDEKRR- 1777

Qy        561 QGGAGRTSPGGRTSPEARGRRSPILLPKGLLAPEAGRADGGTGDSSPSPGSSLPSPLSDP 620
                        |||                                    ::|   | :
Db       1778 -------------LEAR-----------------------------IATLEEELEEE 1792

Qy        621 RREPMNIYNLIAIIRDQIKHLQAAVDRTTELSRQRIASQELGPAVDKDKEALMEEILKLK 680
              :     : :  |  : ||: |     ||||: ::  ||:     :::  ||:|   : |
Db       1793 QSNSEVLLDRAAARQLQIEQL------TTELANEKSNSQK-----NENGRALLER--QNK 1839

Qy        681 SLLSTKREQITTLRTVLKANKQTAEVALANLKSKYENE--------KA------MVTETM 726
              | :   1  |  || :||   | |  :| :: : |||        ||       : |
Db       1840 ELKAKLAEIETAQRTKVKATIATLEAKIAKVEEQLENEGKERLLQQKANRKMDKKIKELT 1899

Qy        727 MKLRNELKALKEDAATFSSLRAMFATRCDEYITQLDEMQRQLAAAEDEKKTLNSLLRMAI 786
              | : :| : : :      |     :|       ||| : :|   : :|:         |
Db       1900 MNIEDERRHVDQHKEQMDKLN----SRIKLLKRNLDETEEELQKEKTQKRKYQRECEDMI 1955

Qy        787 QQKLALTQRLELLELDHEQT-----RRGRAKAAPKTKPA 820
              : : |: : :  |:    :|       |    | :| |
Db       1956 ESQEAMNREINSLKTKLRRTGGIGLSSSRLTGTPSSKRA 1994
```

Search completed: July 1, 2003, 08:31:11
Job time : 21.0815 secs